# **BIGSdb Documentation**

Release 1.12.0

**Keith Jolley** 

November 03, 2015

1	Conc	cepts and terms 3
	1.1	BIGSdb
	1.2	Loci
	1.3	Alleles
	1.5	
	1.5	Profiles
	1.6	Sequence tags
	1.7	Sets
2	BIGS	Sdb dependencies 7
	2.1	Required packages
3	Incto	Illation and configuration of BIGSdb 9
3	3.1	
	3.2	Configuring PostgreSQL
	3.3	Site-specific configuration
	3.4	Setting up the offline job manager
	3.5	Setting up the submission system
	3.6	Periodically delete temporary files
	3.7	Prevent preference database getting too large 13
	3.8	Purging old jobs from the jobs database
	3.9	Log file rotation
	3.10	Upgrading BIGSdb
	3.11	Running the BIGSdb RESTful interface
	5.11	
4		base setup 19
	4.1	Creating databases
	4.2	Database-specific configuration    20
	4.3	XML configuration attributes used in config.xml    20
	4.4	User authentication
	4.5	Setting up the admin user
	4.6	Updating PubMed citations
5	Adm	inistrator's guide 31
0	5.1	Types of user
	5.2	User groups
		0 1
	5.3	Curator permissions
	5.4	Locus and scheme permissions (sequence definition database)

	5.5		36
	5.6		36
	5.7	Setting the first user password	37
	5.8		38
	5.9	Temporarily disabling database updates	38
	5.10	Host mapping	39
	5.11	Improving performance	39
	5.12	Dataset partitioning	41
	5.13		45
	5.14		57
	5.15		50
	5.16		57
	5.17		70
	5.18		76
	5.19		78
	5.20		79
	5.20		31
	5.22		35
	5.22		30 90
			90 93
	5.24		
	5.25	1	96
	5.26		98
	5.27		99
	5.28	Authorizing third-party client software to access authenticated resources	Л
6	Cura	tor's guide 10	)3
	6.1	Adding new sender details	)3
	6.2	Adding new allele sequence definitions	
	6.3	Updating and deleting allele sequence definitions	
	6.4	Retiring allele identifiers	
	6.5	Updating locus descriptions	
	6.6	Adding new scheme profile definitions	
	6.7	Updating and deleting scheme profile definitions	
	6.8	Adding isolate records	
	6.9	Updating and deleting single isolate records	
	6.10	Batch updating multiple isolate records	
	6.11	Deleting multiple isolate records	
	6.12	Uploading sequence contigs linked to isolate records	
		Automated web-based sequence tagging	
	6.13		
	6.14	Projects	
	6.15	Isolate record versioning	)1
7	Cura	ting submitted data 15	55
	7.1	Alleles	56
	7.2	Profiles	53
	7.3	Isolates	70
0	0.001		
8		ne curation tools 17	
	8.1	Automated offline sequence tagging	-
	8.2	Automated offline allele definition	
	8.3	Cleanly interrupting offline curation	-
	8.4	Uploading contigs from the command line	30
9	Defin	ition downloads	21
1	9.1	Allele sequence definitions	
	ו••		• •

	9.2	Scheme profile definitions
10	Data	records 187
	10.1	Isolate records
		Allele definition records
		Sequence tag records
		Profile records
	10.5	Sequence bin records
11	Quer	ying data 195
	11.1	Querying sequences to determine allele identity
	11.2	Searching for specific allele definitions
	11.3	Browsing scheme profile definitions
	11.5	Querying scheme profile definitions
	11.5	Investigating allele differences
	11.6	Browsing isolate data
	11.7	Querying isolate data
	11.8	Querying by allelic profile
		Retrieving isolates by linked publication 231
	11.10	User-configurable options
12		analysis plugins 245
		Locus explorer
	12.2	Field breakdown
	12.3	Two field breakdown
	12.4	Scheme and allele breakdown
	12.5	Sequence bin breakdown
	12.6	Genome comparator
	12.7	BLAST
		BURST
		Codon usage
		Unique combinations
		Polymorphisms
		Presence/absence
	12.13	Tag status         293
13	Data	export plugins 297
		Isolate record export
		Sequence export
		Contig export
	15.5	
14	Subn	itting data using the submission system 307
	14.1	Registering a user account
	14.2	Allele submission
	14.3	Profile submission
	14.4	Isolate submission
	14.5	Removing submissions from your notification list
15	REST	If the second
	15.1	Passing additional/optional parameters
	15.2	Resources
	13.3	Authentication
16	Free	uently asked questions (FAQs) 359
	-	General

	Installation	
17.2	ndix         Query operators         Sequence tag flags         Allele sequence flags	361

365

#### 18 Database schema

#### Gene-by-gene population annotation and analysis

BIGSdb is software designed to store and analyse sequence data for bacterial isolates. Any number of sequences can be linked to isolate records - these can be small contigs assembled from dideoxy sequencing through to whole genomes (complete or multiple contigs generated from parallel sequencing technologies such as 454 or Illumina).

BIGSdb extends the principle of MLST to genomic data, where large numbers of loci can be defined, with alleles assigned by reference to sequence definition databases (which can also be set up with BIGSdb). Loci can also be grouped into schemes so that types can be defined by combinations of allelic profiles, a concept analagous to MLST.

The software has been released under the GNU General Public Licence version 3. The latest version of this documentation can be found at http://bigsdb.readthedocs.org/.

# **Concepts and terms**

# 1.1 BIGSdb

BIGSdb is the software platform - not a specific database. There are many instances of BIGSdb databases, so referring to 'the BIGSdb' is meaningless.

# 1.2 Loci

Loci are regions of the genome that are identified by similarity to a known sequence. They can be defined by DNA or peptide sequence. They are often complete coding sequences (genes), but may represent gene fragments (such as used in MLST), antigenic peptide loops, or indeed any sequence feature.

In versions of BIGSdb prior to 1.8.0, an isolate record could only have one live *allele* designation for a locus (inactive/pending designations could be stored within the database but were unavailable for querying or analysis purposes). Since biology is rarely so clean, and some genomes may contain more than one copy of a gene, later versions of the software allow multiple allele designations for a locus, all of which can be queried and analysed.

Paralogous loci can be difficult to differentiate by sequence similarity alone. Because of this, loci can be further defined by context, where in silico PCR or hybridization reactions can be performed to *filter the genome* to specific regions based on sequence external to the locus.

# **1.3 Alleles**

Alleles are instances of loci. Every unique sequence, either DNA or peptide depending on the locus, is defined as a new allele and these are defined in a sequence definition database, where they are given an allele identifier. These identifiers are usually integers, but can be text strings. Allele identifiers in text format can be constrained by length and formatting.

When a specific allele of a locus is identified within the sequence data of an isolate record, the allele designation, i.e. identifier, is associated with the isolate record. This efficiently stores the sequence variation found within an isolate. Two isolates with the same allele designation for a locus have identical sequences at that locus. Once the sequence variation within a genome has been reduced to a series of allele designations, genomes can be efficiently compared by identifying which loci vary between them.

It is important to note that allele identifiers are usually arbitrary and are allocated sequentially in the order of discovery. Alleles with adjacent identifiers may vary by a single nucleotide or by many.

### 1.4 Schemes

Schemes are collections of loci that may be associated with additional field values. At their simplest they just group loci together. Example uses of simple schemes include:

- · Antibiotic resistance genes
- · Genes involved in specific biochemical pathways
- Antigens
- Vaccine components
- Whole genome MLST (wgMLST)

When schemes are associated with additional fields, one of these fields must be the primary key, i.e. its value uniquely defines a particular combination of alleles at its member loci. The pre-eminent example of this is MLST - where a sequence type (ST) is the primary key that uniquely defines combinations of alleles that make up the MLST profiles. Additional fields can also then be included. The values for these need not be unique. In the MLST example, a field for clonal complex can be included, and the same value for this can be set for multiple STs.

# **1.5 Profiles**

Profiles are instances of *schemes*. A profile consists of a set of *allele identifiers* for the *loci* that comprise the scheme. If the scheme has a primary key field, e.g. sequence type (ST) in MLST schemes, then the unique combination of alleles in a complete profile can be defined by the value of this field.

# **1.6 Sequence tags**

Sequence tags record locus position within an isolate record's sequence bin. The process of creating these tags, is known as *tag-scanning*. A sequence tag consists of:

- sequence bin id this identifies a particular contig
- locus name
- start position
- end position
- flag to indicate if sequence is reversed
- flag to indicate if sequence is complete and does not continue off the end of the contig

# 1.7 Sets

Sets provide a means to take a large database with multiple loci and/or schemes and present a subset of these as though it was a complete database. The loci and schemes chosen to belong to a set can be renamed when used with this set. The rationale for this is that in a database with disparate isolates and a large number of loci, the naming of these loci may have to be long to specify a species name. For example, you may have a database that contains multiple MLST schemes for different species, but since these schemes may use different fragments of the same genes they may have to be named something like 'Streptococcus\_pneumoniae\_MLST\_aroE' to uniquely specify them. If we define a set for 'Streptococcus pneumoniae' we can then choose to only include S. pneumoniae loci and therefore shorten their names, e.g. to 'aroE'.

Additional metadata fields can also be associated with each set so it is possible to have a database containing genomes from multiple species and a generic set of metadata, then have additional specific metadata fields for particular species or genera. These additional fields only become visible and searchable when the specific set that they belong to has been selected.

# **BIGSdb** dependencies

# 2.1 Required packages

BIGSdb requires a number of software components to be installed:

### 2.1.1 Linux packages

- Apache2 web server with mod\_perl2
- PostgreSQL database
- Perl 5.10+
- BioPerl
- BLAST+
- EMBOSS
  - infoalign use to extract alignment stats in Genome Comparator.
  - sixpack used to translate sequences in multiple reading frames.
  - stretcher used for sequence alignment in allele query.
- Ipcress part of exonerate package used to simulate PCR reactions which can be used to filter the genome to predicted amplification products.
- Xvfb X virtual framebuffer needed to support SplitsTree in command line mode as used in Genome Comparator.

### 2.1.2 Perl modules

These are included with most Linux distributions.

- DBI Database independent interface module used to interact with databases.
- DBD-Pg PostgreSQL database driver for DBI.
- XML::Parser::perlSAX part of libxml-perl Used to parse XML configuration files.
- Log::Log4perl Configurable status and error logging.
- Log::Dispatch::File Object for logging to file.
- Error Exception handling.

- Config::Tiny Configuration file handling.
- Bio::Biblio This used to be part of BioPerl but will need to be installed separately if using BioPerl 1.6.920 or later.
- IO::String
- Data::UUID Globally unique identifer handling for preference storage.
- List::MoreUtils (version 0.28+).
- Time::Duration [optional] Used by Job Viewer to display elapsed time in rounded units.
- Excel::Writer::XLSX Used to export data in Excel format.
- Parallel::ForkManager Required for multi-threading autotagger and autodefiner scripts.
- Net::Oauth Required for REST authentication (this needs to be installed even if you are not using REST).
- Crypt::Eksblowfish::Bcrypt Used for password hashing.
- Mail::Sender [optional] Used to send E-mail messages by submission system.

### 2.1.3 Optional packages

Installing these packages will enable extra functionality, but they are not required by the core BIGSdb package.

- ChartDirector library used for generating charts. Used by some plugins.
- ImageMagick mogrify used by some plugins.
- MAFFT 6.8+ sequence alignment used by some plugins.
- Muscle sequence alignment used by some plugins.
- Splitstree4 used by GenomeComparator plugin.

# Installation and configuration of BIGSdb

# 3.1 Software installation

BIGSdb consists of two main Perl scripts, bigsdb.pl and bigscurate.pl, that run the query and curator's interfaces respectively. These need to be located somewhere within the web cgi-bin directories. In addition, there are a large number of library files, used by both these scripts, that are installed by default in /usr/local/lib/BIGSdb. Plugin scripts are stored within a 'Plugins' sub-directory of this library directory.

All databases on a system can use the same instance of the scripts, or alternatively any database can specify a particular path for each script, enabling these script directories to be protected by apache htaccess directives.

- Software requirements
- Download from SourceForge.net or GitHub.
- 1. Unpack the distribution package in a temporary directory:

```
gunzip bigsdb_1.x.x.tar.gz
tar xvf bigsdb_1.x.x.tar
```

- 2. Copy the bigsdb.pl and bigscurate.pl scripts to a subdirectory of your web server's cgi-bin directory. Make sure these are readable and executable by the web server daemon.
- Copy the contents of the lib directory to /usr/local/lib/BIGSdb/. Make sure you include the Plugins and Offline directories which are subdirectories of the main lib directory.
- Copy the contents of the javascript directory to a javascript directory within the web root tree, i.e. accessible from http://your\_website/javascript/.
- 5. Copy the contents of the css directory to a css directory within the web root tree, i.e. accessible from <a href="http://your\_website/css/">http://your\_website/css/</a>.
- 6. Copy the contents of the fonts directory to a fonts directory within the web root tree, i.e. accessible from <a href="http://your\_website/fonts/">http://your\_website/fonts/</a>.
- 7. Copy the images directory to the root directory of your website, i.e. accessible from http://your\_website/images/.
- 8. Copy the contents of the conf directory to /etc/bigsdb/. Check the paths of helper applications and database names in the bigsdb.conf file and modify for your system.
- 9. Create a PostgreSQL database user called apache this should not have any special priveleges. First you will need to log in as the postgres user:

sudo su postgres

Then use the createuser command to do this, e.g.

createuser apache

From the psql command line, set the apache user password:

```
psql
ALTER ROLE apache WITH PASSWORD 'remote';
```

10. Create PostgreSQL databases called bigsdb\_auth, bigsdb\_prefs and bigsdb\_refs using the scripts in the sql directory. Create the database using the createdb command and set up the tables using the psql command.

```
createdb bigsdb_auth
psql -f auth.sql bigsdb_auth
createdb bigsdb_prefs
psql -f prefs.sql bigsdb_prefs
createdb bigsdb_refs
psql -f refs.sql bigsdb_refs
```

- 11. Create a writable temporary directory in the root of the web site called tmp, i.e. accessible from http://your\_website/tmp.
- 12. Create a log file, bigsdb.log, in /var/log owned by the web server daemon, e.g.

```
touch /var/log/bigsdb.log
chown www-data /var/log/bigsdb.log
```

(substitute www-data for the web daemon user).

# 3.2 Configuring PostgreSQL

PostgreSQL can be configured in many ways and how you do this will depend on your site requirements.

The following security settings will allow the appropriate users 'apache' and 'bigsdb' to access databases without allowing all logged in users full access. Only the UNIX users 'postgres' and 'webmaster' can log in to the databases as the Postgres user 'postgres'.

You will need to edit the pg\_hba.conf and pg\_ident.conf files. These are found somewhere like /etc/postgresql/9.1/main/

#### pg\_hba.conf

```
# Database administrative login by UNIX sockets
       all
local
                   postgres
                                                      ident map=mymap
# TYPE DATABASE
                   USER
                                CIDR-ADDRESS
                                                      METHOD
# "local" is for Unix domain socket connections only
local all
                 all
                                                      ident map=mymap
# IPv4 local connections:
                                127.0.0.1/32
                                                      md5
host
       all
                   all
# IPv6 local connections:
                   all
                                                     md5
host
       all
                                ::1/128
```

#### pg\_ident.conf

# MAPNAME	SYSTEM-USERNAME	PG-USERNAME
тутар	postgres	postgres
тутар	webmaster	postgres
тутар	www-data	apache

mymap	bigsdb	bigsdb	
mymap	bigsdb	apache	

You may also need to change some settings in the postgresql.conf file. As an example, a configuration for a machine with 16GB RAM, allowing connections from a separate web server may have the following configuration changes made:

```
listen_addresses = '*'
max_connections = 200
shared_buffers = 1024Mb
work_mem = 8Mb
effective_cache_size = 8192Mb
stats_temp_directory = '/dev/shm'
```

Setting stats\_temp\_directory to /dev/shm makes use of a ramdisk usually available on Debian or Ubuntu systems for frequently updated working files. This reduces a lot of unneccessary disk access.

See Tuning Your PostgreSQL Server for more details.

Restart PostgreSQL after any changes, e.g.

/etc/init.d/postgresql restart

### 3.3 Site-specific configuration

Site-specific configuration files are located in /etc/bigsdb by default.

- bigsdb.conf main configuration file
- logging.conf error logging settings. See log4perl project website for advanced configuration details.

# 3.4 Setting up the offline job manager

To run plugins that require a long time to complete their analyses, an offline job manager has been developed. The plugin will save the parameters of a job to a job database and then provide a link to the job status page. An offline script, run frequently from CRON, will then process the job queue and update status and outputs via the job status page.

1. Create a 'bigsdb' UNIX user, e.g.:

sudo useradd -s /bin/sh bigsdb

2. As the postgres user, create a 'bigsdb' user and create a bigsdb\_jobs database using the jobs.sql SQL file, e.g.:

```
createuser bigsdb [no need for special priveleges]
createdb bigsdb_jobs
psql -f jobs.sql bigsdb_jobs
```

From the psql command line, set the bigsdb user password::

psql ALTER ROLE bigsdb WITH PASSWORD 'bigsdb';

3. Set up the jobs parameters in the /etc/bigsdb/bigsdb.conf file, e.g.:

```
jobs_db=bigsdb_jobs
max_load=8
```

The jobs script will not process a job if the server's load average (over the last minute) is higher than the max\_load parameter. This should be set higher than the number of processor cores or you may find that jobs never run on a busy server. Setting it to double the number of cores is probably a good starting point.

- 4. Copy the job\_logging.conf file to the /etc/bigsdb directory.
- 5. Set the script to run frequently (preferably every minute) from CRON. Note that CRON does not like '.' in executable filenames, so either rename the script to 'bigsjobs' or create a symlink and call that from CRON, e.g.:

```
copy bigsjobs.pl to /usr/local/bin
sudo ln -s /usr/local/bin/bigsjobs.pl /usr/local/bin/bigsjobs
```

You should install xvfb, which is a virtual X server that may be required for third party applications called from plugins. This is required, for example, for calling splitstree4 from the Genome Comparator plugin.

Add the following to /etc/crontab::

\* \* \* \* \* bigsdb xvfb-run -a /usr/local/bin/bigsjobs

(set to run every minute from the 'bigsdb' user account).

If you'd like to run this more frequently, e.g. every 30 seconds, multiple entries can be added to CRON with an appropriate sleep prior to running, e.g.:

```
* * * * bigsdb xvfb-run -a /usr/local/bin/bigsjobs
* * * * * bigsdb sleep 30;xvfb-run -a /usr/local/bin/bigsjobs
```

6. Create a log file, bigsdb\_jobs.log, in /var/log owned by 'bigsdb', e.g.:

```
sudo touch /var/log/bigsdb_jobs.log
sudo chown bigsdb /var/log/bigsdb_jobs.log
```

# 3.5 Setting up the submission system

The submission system allows users to submit new data to the database for curation. Submissions are placed in a queue for a curator to upload. All communication between submitters and curators can occur via the submission system.

 Create a writable submissions directory in the root of the web site called submissions, i.e. accessible from http://your\_website/submissions. This is used for file uploads. The directory should be writable by the Apache web daemon (user 'www-data' on Debian/Ubuntu systems). If you are running the *RESTful interface* the directory should also be writable by the bigsdb user. To ensure this, make the directory group-writable and add the bigsdb user to the apache group ('www-data' on Debian/Ubuntu systems). If you will be allowing submissions via the RESTful interface, you should also add the apache user ('www-data' on Debian/Ubuntu systems) to the bigsdb group, e.g.

```
sudo usermod -a -G www-data bigsdb
sudo usermod -a -G bigsdb www-data
```

The actual directory can be outside of the web root and made accessible using a symlink provided your Apache configuration allows this, e.g. the default location is /var/submissions symlinked to /var/www/submissions (assuming your web site is located in /var/www), e.g.

```
sudo touch /var/submissions
sudo chown www-data:www-data /var/submissions
sudo chmod 775 /var/submissions
sudo ln -s /var/submissions /var/www
```

- 2. Set the submission\_dir location in bigsdb.conf.
- 3. Set the smtp\_server in bigsdb.conf to the IP or DNS name of your organisation's SMTP relay. Depending on how your E-mail system is configured, you may be able to use the localhost address (127.0.0.1).
- 4. Make sure the curate\_script and query\_script values are set in bigsdb.conf. These point to the web-accessible location of the web scripts and are required to allow curators to be directed between the web interfaces as needed.
- 5. Set submissions="yes" in the system tag of the *database config.xml file* of each database for which submissions should be enabled.

# 3.6 Periodically delete temporary files

There are two temporary directories (one public, one private) which may accumulate temporary files over time. Some of these are deleted automatically when no longer required but some cannot be cleaned automatically since they are used to display results after clicking a link or to pass the database query between pages of results.

The easiest way to clean the temp directories is to run a cleaning script periodically, e.g. create a root-executable script in /etc/cron.hourly containing the following::

```
#!/bin/sh
#Remove temp BIGSdb files from secure tmp folder older than 1 week.
find /var/tmp/ -name '*BIGSdb_*' -type f -mmin +10080 -exec rm -f {} \; 2>/dev/null
#Remove .jnlp files from web tree older than 1 day
find /var/www/tmp/ -name '*.jnlp' -type f -mmin +1440 -exec rm -f {} \; 2>/dev/null
#Remove other tmp files from web tree older than 1 week
find /var/www/tmp/ -type f -mmin +10080 -exec rm -f {} \; 2>/dev/null
```

# 3.7 Prevent preference database getting too large

The preferences database stores user preferences for BIGSdb databases running on the site. Every user will have a globally unique identifier (guid) stored in this database along with a datestamp indicating the last access time. On public databases that do not require logging in, this guid is stored as a cookie on the user's computer. Databases that require logging in use a combination of database and username as the identifier. Over time, the preferences database can get quite large since every unique user will result in an entry in the database. Since many of these entries represent casual users, or even web indexing bots, they can be periodically cleaned out based on their last access time. A weekly CRON job can be set up to remove any entries older than a defined period. For example, the following line entered in /etc/crontab will remove the preferences for any user that has not accessed any database in the past 6 months (the script will run at 6pm every Sunday).

#Prevent prefs database getting too large
00 18 \* \* 0 postgres psql -c "DELETE FROM guid WHERE last\_accessed < NOW() - INTERVAL '6 monthlands'</pre>

### 3.8 Purging old jobs from the jobs database

If you are running the offline job manager, the jobs database (default bigsdb\_jobs) contains the parameters and output messages of these jobs. Job output files are only *usually kept on the server for 7 days* so there is no point keeping the database entries for longer than this. These can be purged with a daily cron job, e.g. set the following in /etc/crontab (the script will run at 5am every day).

```
#Purge jobs older than 7 days from the jobs database.
00 5 * * * postgres psql -c "DELETE FROM jobs where (stop_time IS NOT NULL AND stop_time < now</pre>
```

# 3.9 Log file rotation

Set the log file to auto rotate by adding a file called 'bigsdb' with the following contents to /etc/logrotate.d:

```
/var/log/bigsdb.log {
 weeklv
 rotate 4
 compress
 copytruncate
 missingok
 notifempty
 create 640 root adm
/var/log/bigsdb_jobs.log {
 weekly
 rotate 4
 compress
 copytruncate
 missingok
 notifempty
 create 640 root adm
```

### 3.10 Upgrading BIGSdb

Major version changes, e.g.  $1.7 \rightarrow 1.8$ , indicate that there has been a change to the underlying database structure for one or more of the database types. Scripts to upgrade the database are provided in sql/upgrade and are named by the database type and version number. For example, to upgrade an isolate database (bigsdb\_isolates) from version 1.7 to 1.8, log in as the postgres user and type:

```
psql -f isolatedb_v1.8.sql bigsdb_isolates
```

Upgrades are sequential, so to upgrade from a version earlier than the last major version you would need to upgrade to the intermediate version first, e.g. to go from  $1.6 \rightarrow 1.8$ , requires upgrading to 1.7 first.

Minor version changes, e.g. 1.8.0 -> 1.8.1, have no modifications to the database structures. There will be changes to the Perl library modules and possibly to the contents of the Javascript directory, images directory and CSS files. The version number is stored with the bigsdb.pl script, so this should also be updated so that BIGSdb correctly reports its version.

# 3.11 Running the BIGSdb RESTful interface

BIGSdb has an Application Programming Interface (API) that allows third-party applications to access the data within the databases. The script that runs this is called bigsrest.pl. This is a Dancer2 application that can be run using a wide range of options, e.g. as a stand-alone script, using Perl webservers with plackup, or from apache. Full documentation for deploying Dancer2 applications can be found online.

The script requires a new database that describes the resources to make available. This is specified in the bigsdb.conf file as the value of the 'rest\_db' attribute. By default, the database is named bigsdb\_rest.

A SQL file to create this database can be found in the sql directory of the download archive. It is called rest.sql. To create the database, as the postgres user, navigate to the sql directory and type

```
createdb bigsdb_rest
psql -f rest.sql bigsdb_rest
```

This database will need to be populated using psql or any tool that can be used to edit PostgreSQL databases. The database contains three tables that together describe and group the databases resources that will be made available through the API. The tables are:

#### resources

- this contains two fields (both compulsory):
  - \* **dbase\_config** the name of the database configuration used with the database. This is the same as the name of the directory that contains the config.xml file in the /etc/bigsdb/dbases directory.
  - \* description short description of the database.
- groups (used to group related resources together)
  - this contains two fields (compulsory fields shown in bold):
    - \* **name** short name of group. This is usually a single word and is also the key that links resources to groups.
    - \* **description** short description of group.
    - \* long\_description fuller description of group.
- group\_resources (used to add resources to groups)
  - this contains two fields (both compulsory)
    - \* group\_name name of group. This must already exist in the groups table.
    - \* dbase\_config the name of database resource. This must already exist in the resources table.

For example, to describe the PubMLST resources for Neisseria, connect to the bigsdb\_rest database using psql,

psql bigsdb\_rest

Then enter the following SQL commands. First add the database resources:

```
INSERT INTO resources (dbase_config,description) VALUES
('pubmlst_neisseria_seqdef','Neisseria sequence/profile definitions');
INSERT INTO resources (dbase_config,description) VALUES
('pubmlst_neisseria_isolates','Neisseria isolates');
```

Then create a 'neisseria' group that will contain these resources:

```
INSERT INTO groups (name,description) VALUES
('neisseria','Neisseria spp.');
```

Finally, add the database resources to the group:

```
INSERT INTO group_resources (group_name,dbase_config) VALUES
('neisseria','pubmlst_neisseria_seqdef');
INSERT INTO group_resources (group_name,dbase_config) VALUES
('neisseria','pubmlst_neisseria_isolates');
```

The REST API will need to run on its own network port. By default this is port 3000. To run as a stand-alone script, from the script directory, as the bigsdb user, simply type:

./bigsrest.pl

This will start the API on port 3000. You will be able to check that this is running using a web browser by navigating to http://localhost:3000 on the local machine, or using the server IP address from a remote machine. You may need to modify your server firewall rules to allow connection to this port.

Running as a stand-alone script is useful for testing, but you can achieve much better performance using a Perl webserver with plackup. There are various options to choose. PubMLST uses Starman.

To run the API using Starman, type the following as the bigsdb user:

plackup -a /var/rest/bigsrest.pl -s Starman -E deployment

where the value of -a refers to the location of the bigsrest.pl script. Starman defaults to using port 5000.

Different Linux distributions use different means to control services/daemons. To start the REST interface on system boot on systems using upstart, create a file called bigsdb-rest.conf in /etc/init. The contents of this file should be something like (modify file paths as appropriate):

```
description "Start BIGSdb REST interface"
version "1.0"
author "Keith Jolley"
start on runlevel [12345]
## tell upstart we're creating a daemon
expect fork
script
exec su -s /bin/sh -c 'exec "$0" "$@"' bigsdb -- /usr/local/bin/plackup -a /var/rest/bigsrest.pl -s send script
```

#### 3.11.1 Proxying the API to use a standard web port

Usually you will want your API to be available on the standard web port 80. To do this you will need to set up a virtual host using a different domain name from your web site to proxy the API port. For example, PubMLST has a separate domain 'http://rest.pubmlst.org' for its API. This is set up as a virtual host directive in apache with the following configuration file:

```
<VirtualHost *>
ServerName rest.pubmlst.org
DocumentRoot /var/rest
ServerAdmin keith.jolley@zoo.ox.ac.uk
<Directory /var/rest>
AllowOverride None
Require all granted
</Directory>
```

```
ProxyPass / http://rest.pubmlst.org:5000/
ProxyPassReverse / http://rest.pubmlst.org:5000/
</Proxy *>
    Order allow,deny
    Allow from all
</Proxy>
ErrorLog /var/log/apache2/rest.pubmlst.org-error.log
CustomLog /var/log/apache2/rest.pubmlst.org-access.log common
</VirtualHost>
```

### **Database setup**

There are two types of BIGSdb database:

- · sequence definition databases, containing
  - allele sequences and their identifiers
  - scheme data, e.g. MLST profile definitions
- isolate databases, containing
  - isolate provenance metadata
  - genome sequences
  - allele designations for loci defined in sequence definition databases.

These two databases are independent but linked. A single isolate database can communicate with multiple sequence definition databases and vice versa. Different access restrictions can be placed on different databases.

Databases are described in XML files telling BIGSdb everything it needs to know about them. Isolate databases can have any fields defined for the isolate table, allowing customisation of metadata - these fields are described in the XML file (config.xml) and must match the fields defined in the database itself.

### 4.1 Creating databases

There are templates available for the sequence definition and isolate databases. These are SQL scripts found in the sql directory.

To create a database, you will need to log in as the postgres user and use these templates. For example to create a new sequence definition database called bigsdb\_test\_sequef, navigate to the sql directory and log in as the postgres user, e.g.:

```
sudo su postgres
```

then

```
createdb bigsdb_test_seqdef
psql -f seqdef.sql bigsdb_test_seqdef
```

#### Create an isolate database the same way:

```
createdb bigsdb_test_isolates
psql -f isolatedb.sql bigsdb_test_isolates
```

The standard fields in the isolate table are limited to essential fields required by the system. To add new fields, you need to log in to the database and alter this table. For example, to add fields for country and year, first log in to the newly created isolate database as the postgres user::

psql bigsdb\_test\_isolates

and alter the isolate table::

```
ALTER TABLE isolates ADD country text;
ALTER TABLE isolates ADD year int;
```

Remember that any fields added to the table need to be described in the config.xml file for this database.

# 4.2 Database-specific configuration

Each BIGSdb database on a system has its own configuration directory, by default in /etc/bigsdb/dbases. The database has a short configuration name used to specify it in a web query and this matches the name of the configuration sub-directory, e.g. http://pubmlst.org/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst\_neisseria\_isolates is the URL of the front page of the PubMLST Neisseria isolate database whose configuration settings are stored in /etc/bigsdb/dbases/pubmlst\_neisseria\_isolates. This database sub-directory contains a number of files (hyperlinks lead to the files used on the Neisseria database):

- config.xml the database configuration file. Fields defined here correspond to fields in the isolate table of the database.
- banner.html optional file containing text that will appear as a banner within the database index pages. HTML markup can be used within this text.
- header.html HTML markup that is inserted at the top of all pages. This can be used to set up site-specific menubars and logos.
- footer.html HTML markup that is inserted at the bottom of all pages.
- curate\_header.html HTML markup that is inserted at the top of all curator's interface pages.
- curate\_footer.html HTML markup that is inserted at the bottom of all curator's interface pages.

### 4.3 XML configuration attributes used in config.xml

The following lists describes the attributes used in the config.xml file that is used to describe databases.

#### 4.3.1 Isolate database XML attributes

Please note that database structure described by the field and sample elements must match the physical structure of the database isolate and sample tables respectively. Required attributes are in **bold**:

<db>

Top level element. Contains child elements: system, field and sample .:

<system>

#### • authentication

- Method of authentication: either 'builtin' or 'apache'. See *user authentication*.

#### • db

- Name of database on system.
- dbtype
  - Type of database: either 'isolates' or 'sequences'.
- description
  - Description of database used throughout interface.
- align\_limit
  - Overrides the sequence export record alignment limit in the Sequence Export plugin. Default: '200'.
- all\_plugins
  - Enable all appropriate plugins for database: either 'yes' or 'no', default 'no'.
- annotation
  - Semi-colon separated list of accession numbers with descriptions (separated by a l), eg. 'AL157959/Z2491;AM421808/FAM18;NC\_002946/FA 1090;NC\_011035/NCCP11945;NC\_014752/020-06'. Currently used only by Genome Comparator plugin.
- cache\_schemes
  - Enable automatic refreshing of scheme field caches when batch adding new isolates: either 'yes' or 'no', default 'no'.
  - See scheme caching.
- codon\_usage\_limit
  - Overrides the record limit for the Codon Usage plugin. Default: '500'.
- contig\_analysis\_limit
  - Overrides the isolate number limit for the Contig Export plugin. Default: '1000'.
- curate\_only
  - Set to 'yes' to hide field on an isolate information page in the standard interface. The field will be visible if the page is accessed via the curator's interface (version 1.10.0+).
- curate\_path\_includes
  - Partial path of the bigscurate.pl script used to curate the database. See user authentication.
- curate\_script
  - Relative web path to curation script. Default 'bigscurate.pl' (version 1.11+).
  - This is only needed if automated submissions are enabled. If bigscurate.pl is in a different directory from bigsdb.pl, you need to include the whole web path, e.g. /cgi-bin/private/bigsdb/bigscurate.pl.
- default
  - Default value. This will be entered automatically in the web form but can be overridden.
- default\_access
  - The default access to the database configuration, either 'allow' or 'deny'. If 'allow', then specific users can be denied access by creating a file called 'users.deny' containing usernames (one per line) in the configuration directory. If 'deny' then specific users can be allowed by creating a file called 'users.allow' containing usernames (one per line) in the configuration directory. See *default access*.
- default\_seqdef\_config

- Isolate databases only: Name of the default seqdef database configuration used with this database. Used to automatically fill in details when adding new loci.
- default\_seqdef\_dbase
  - Isolate databases only: Name of the default seqdef database used with this database. Used to automatically fill in details when adding new loci.
- default\_seqdef\_script
  - Isolate databases only: URL of BIGSdb script running the seqdef database (default: '/cgibin/bigsdb/bigsdb.pl').
- fieldgroup1 fieldgroup10
  - Allows multiple fields to be queried as a group. Value should be the name of the group followed by a colon
    (:) followed by a comma-separated list of fields to group, e.g. identifiers:id,strain,other\_name.
- genome\_comparator\_limit
  - Overrides the isolate number limit for the Genome Comparator plugin. Default: '1000'.
- genome\_comparator\_max\_ref\_loci
  - Overrides the limit on number of loci allowed in a reference genome. Default: '10000'.
- hide\_unused\_schemes
  - Sets whether a scheme is shown in a main results table if none of the isolates on that page have any data for the specific scheme: either 'yes' or 'no', default 'no'.
- host
  - Host name/IP address of machine hosting isolate database, default 'localhost'.
- job\_priority
  - Integer with default job priority for offline jobs (default:5).
- job\_quota
  - Integer with number of offline jobs that can be queued or currently running for this database.
- labelfield
  - Field that is used to describe record in isolate info page, default 'isolate'.
- locus\_aliases
  - Display locus aliases and use them in dropdown lists by default: must be either 'yes' or 'no', default 'no'. This option can be overridden by a user preference.
- locus\_superscript\_prefix
  - Superscript the first letter of a locus name if it is immediately following by an underscore, e.g. f\_abcZ would be displayed as fabcZ within the interface: must be either 'yes' or 'no', default 'no'. This can be used to designate gene fragments (or any other meaning you like).
- maindisplay\_aliases
  - Default setting for whether isolates aliases are displayed in main results tables: either 'yes' or 'no', default 'no'. This setting can be overridden by individual user preferences.
- noshow
  - Comma-separated list of fields not to use in breakdown statistic plugins.
- no\_publication\_filter

- Isolate databases only: Switches off display of publication filter in isolate query form by default: either 'yes' or 'no', default 'no'.
- only\_sets
  - Don't allow option to view the 'whole database' only list sets that have been defined: either 'yes' or 'no', default 'no'.
- password
  - Password for access to isolates database, default 'remote'.
- port
  - Port number that the isolate host is listening on, default '5432'.
- privacy
  - Displays E-mail address for sender in isolate information page if set to 'no'. Default 'yes'.
- query\_script
  - Relative web path to bigsdb script. Default 'bigsdb.pl' (version 1.11+).
  - This is only needed if automated submissions are enabled. If bigsdb.pl is in a different directory from bigscurate.pl, you need to include the whole web path, e.g. /cgi-bin/bigsdb/bigsdb.pl.
- read\_access
  - Describes who can view data: either 'public' for everybody or 'authenticated\_users' for anybody who has been able to log in. Default 'public'.
- script\_path\_includes
  - Partial path of the bigsdb.pl script used to access the database. See user authentication.
- seqbin\_size\_threshold
  - Sets the size values in Mbp to enable for the seqbin filter.
  - Example: seqbin\_size\_threshold="0.5,1,2,4".
- seq\_export\_limit
  - Overrides the sequence export limit (records x loci) in the Sequence Export plugin. Default: '1000000'.
- sets
  - Use sets: either 'yes' or 'no', default 'no'.
- set\_id
  - Force the use of a specific set when accessing database via this XML configuration: Value is the name of the set.
- start\_id
  - Defines the minimum record id to be used when uploading new isolate records. This can be useful when it is anticipated that two databases may be merged and it would be easier to do so if the id numbers in the two databases were different. Default: '1'.
- submissions
  - Enable automated submission system: either 'yes' or 'no', default 'no' (version 1.11+).
  - The curate\_script and query\_script paths should also be set, either in the bigsdb.conf file (for site-wide configuration) or within the system attribute of config.xml.
- tblastx\_tagging

- Sets whether tagging can be performed using TBLASTX: either 'yes' or 'no', default 'no'.

• user

- Username for access to isolates database, default 'apache'.
- use\_temp\_scheme\_table
  - Sets whether entire schemes are imported in to the isolate database in to an indexed table rather than querying the seqdef scheme view for isolate results tables. Under some circumstances this can be considerably quicker than querying the seqdef scheme view (a few ms compared to >10s if the seqdef database contains multiple schemes with an uneven distribution of a large number of profiles so that the Postgres query planner picks a sequential rather than index scan). This scheme table can also be generated periodically using the update\_scheme\_cache.pl script to create a persistent cache. This is particularly useful for large schemes (>10000 profiles) but data will only be as fresh as the cache so ensure that the update script is run periodically.
- view
  - Database view containing isolate data, default 'isolates'.
- views
  - Comma-separated list of views of the isolate table defined in the database. This is used to set a view for a set.
- webroot
  - URL of web root, which can be relative or absolute. The bigsdb.css stylesheet file should be located in this directory. Default '/'.

<field>

Element content: Field name + optional list <optlist> of allowed values, e.g.:

```
<field type="text" required="no" length="40" maindisplay="no"

web="http://somewebsite.com/cgi-bin/script.pl?id=[?]" optlist="yes">epidemiology

<optlist>

<option>carrier</option>

<option>healthy contact</option>

<option>sporadic case</option>

<option>endemic</option>

<option>epidemic</option>

<option>pandemic</option>

</optlist>

</field>
```

• type

- Data type: int, text, float, bool, or date.
- comments \* optional
  - Comments about the field. These will be displayed in the field description plugin and as tooltips within the curation interface.
- dropdown
  - Select if you want this field to have its own dropdown filter box on the query page. If the field has an option list it will use the values in it, otherwise all values defined in the database will be included: 'yes' or 'no', default 'no'. This setting can be overridden by individual user preferences.
- length
  - Length of field, default 12.

- maindisplay
  - Sets if field is displayed in the main table after a database search, 'yes' or 'no', default 'yes'. This setting can be overridden by individual user preferences.
- max
  - Maximum value for integer types. Special values such as CURRENT\_YEAR can be used.
- min
  - Minimum value for integer types.
- optlist
  - Sets if this field has a list of allowed values, default 'no'. Surround each option with an <option> tag.
- regex
  - Regular expression used to constrain field values, e.g. regex="^[A-Z].\*\$" forces the first letter of the value to be capitalized.
- required
  - Sets if data is required for this field, 'yes' or 'no', default 'yes'.
- userfield
  - Select if you want this field to have its own dropdown filter box of users (populated from the users table): 'yes' or 'no', default 'no'.

• web

- URL that will be used to hyperlink field values. If [?] is included in the URL, this will be substituted for the actual field value.

#### **Special values**

The following special variables can be used in place of an actual value:

CURRENT\_YEAR: the 4 digit value of the current year

<sample>

Element content: Sample field name + optional list <optlist> of allowed values. Attributes are essentially the same as isolate field attributes, but refer to the samples table rather than the isolates table.

The sample table, if defined, must include isolate\_id and sample\_id fields, which must also be described in the XML file. These must be set as integer fields.

#### 4.3.2 Sequence definition database XML attributes

Required attributes are in **bold**.

<db>

Top level element. Contains child elements: system, field and sample.

<system>

- authentication
  - Method of authentication: either 'builtin' or 'apache'. See user authentication.

#### • db

- Name of database on system.
- dbtype
  - Type of database: either 'isolates' or 'sequences'.
- description
  - Description of database used throughout interface.
- align\_limit
  - Overrides the sequence export record alignment limit in the Sequence Export plugin. Default: '200'.
- allele\_comments
  - Enable comments on allele sequences: either 'yes' or 'no', default 'no'.
  - This is not enabled by default to discourage the practice of adding isolate information to allele definitions (this sort of information belongs in an isolate database).
- allele\_flags
  - Enable flags to be set for alleles: either 'yes' or 'no', default 'no'.
- curate\_path\_includes
  - Partial path of the bigscurate.pl script used to curate the database. See user authentication.
- curate\_script
  - Relative web path to curation script. Default 'bigscurate.pl' (version 1.11+).
  - This is only needed if automated submissions are enabled. If bigscurate.pl is in a different directory from bigsdb.pl, you need to include the whole web path, e.g. /cgi-bin/private/bigsdb/bigscurate.pl.
- · diploid
  - Allow IUPAC 2-nuclotide ambiguity codes in allele definitions for use with diploid typing schemes: either 'yes' or 'no', default 'no'.
- disable\_seq\_downloads
  - Prevent users or curators from downloading all alleles for a locus (admins always can). 'yes' or 'no', default 'no'.
- job\_priority
  - Integer with default job priority for offline jobs (default:5).
- job\_quota
  - Integer with number of offline jobs that can be queued or currently running for this database.
- materialized\_views
  - Enable materialized views: either 'yes' or 'no', default 'no'.
- profile\_submissions
  - Enable profile submissions (automated submission system): either 'yes' or 'no', default 'no' (version 1.11+).
  - To enable, you will also need to set submissions="yes". By default, profile submissions are disabled since generally new profiles should be accompanied by representative isolate data, and the profile can be extracted from that.

- query\_script
  - Relative web path to bigsdb script. Default 'bigsdb.pl' (version 1.11+).
  - This is only needed if automated submissions are enabled. If bigsdb.pl is in a different directory from bigscurate.pl, you need to include the whole web path, e.g. /cgi-bin/bigsdb/bigsdb.pl.
- read\_access
  - Describes who can view data: either 'public' for everybody, or 'authenticated\_users' for anybody who has been able to log in. Default 'public'.
- script\_path\_includes
  - Partial path of the bigsdb.pl script used to access the database. See user authentication.
- seq\_export\_limit
  - Overrides the sequence export limit (records x loci) in the Sequence Export plugin. Default: '1000000'.
- sets
  - Use sets: either 'yes' or 'no', default 'no'.
- set\_id
  - Force the use of a specific set when accessing database via this XML configuration: Value is the name of the set.
- submissions
  - Enable automated submission system: either 'yes' or 'no', default 'no' (version 1.11+).
  - The curate\_script and query\_script paths should also be set, either in the bigsdb.conf file (for site-wide configuration) or within the system attribute of config.xml.
- webroot
  - URL of web root, which can be relative or absolute. The bigsdb.css stylesheet file should be located in this directory. Default '/'.

### 4.4 User authentication

You can choose whether to allow Apache to handle your authentication or use built-in authentication.

### 4.4.1 Apache authentication

Using apache to provide your authentication allows a flexible range of methods and back-ends (see the Apache authentication HowTo for a start, or any number of tutorials on the web).

At its simplest, use a .htaccess file in the directory containing the bigscurate.pl (and bigsdb.pl for restriction of readaccess) script or by equivalent protection of the directory in the main Apache server configuration. It is important to note however that, by default, any BIGSdb database can be accessed by any instance of the BIGSdb script (including one which may not be protected by a .htaccess file, allowing public access). To ensure that only a particular instance (protected by a specific htaccess directive) can access the database, the following attributes can be set in the system tag of the database XML description file:

- script\_path\_includes: the BIGSdb script path must contain the value set.
- curate\_path\_includes: the BIGSdb curation script path must contain the value set.

For public databases, the 'script\_path\_includes' attribute need not be set.

To use apache authentication you need to set the authentication attribute in the system tag of the database XML configuration to 'apache'.

### 4.4.2 Built-in authentication

BIGSdb has its own built-in authentication, using a separate database to store password and session hashes. The advantages of using this over many forms of apache authentication are:

- Users are able to update their own passwords.
- Passwords are not transmitted over the Internet in plain text.

When a user logs in, the server provides a random one-time session variable and the user is prompted to enter their username and password. The password is encrypted within the browser using a Javscript one-way hash algorithm, and this is combined with the session variable and hashed again. This hash is passed to the server. The server compares this hash with its own calculated hash of the stored encrypted password and session variable that it originally sent to the browser. Implementation is based on perl-md5-login.

To use built-in authentication you need to set the authentication attribute in the system tag of the database XML configuration to 'builtin'.

# 4.5 Setting up the admin user

The first admin user needs to be manually added to the users table of the database. Connect to the database using psql and add the following (changing details to suit the user).:

```
INSERT INTO users (id, user_name, surname, first_name, email, affiliation, status, date_entered,
datestamp, curator) VALUES (1, 'keith', 'Jolley', 'Keith', 'keith.jolley@zoo.ox.ac.uk',
'University of Oxford, UK', 'admin', 'now', 'now', 1);
```

If you are using built-in authentication, set the password for this user using the *add\_user.pl* script. This hashes the password and stores this within the authentication database. Other users can be added by the admin user from the curation interface accessible from http://your\_website/cgi-bin/private/bigscurate.pl?db=test\_db (or wherever you have located your bigscurate.pl script).

# 4.6 Updating PubMed citations

Publications listed in PubMed can be associated with individual isolate records, profiles, loci and sequences. Full citations for these are stored within a local reference database, enabling these to be displayed within isolate records and searching by publication and author. This local database is populated by a script that looks in BIGSdb databases for PubMed records not locally stored and then requests the full citation record from the PubMed database.

The script is called getrefs.pl and can be found in the scripts/maintenance directory. This script needs to know which BIGSdb databases and tables it needs to search for PubMed ids. These are listed in a configuration file (usually called getrefs.conf) which contains two columns - the first is the name of the database, the second is a comma-separated list of tables to search, e.g.

```
pubmlst_bigsdb_neisseria_isolatesrefspubmlst_bigsdb_neisseria_seqdefprofile_refs, sequence_refs, locus_refs
```

The script can be called as follows:

getrefs.pl getrefs.conf

Run either as the 'postgres' user or an account that is allowed to connect as the postgres user.

This should be run periodically from a CRON job, e.g. every hour.

## Administrator's guide

Please note that links displayed within the curation interface will vary depending on database contents and the permissions of the curator.

## 5.1 Types of user

There are four types of user in BIGSdb:

- User can view data but never modify it. Users should be created for every submitter of data so that records can be tracked, even if they do not actually use the database.
- Submitter (isolate databases only) can add and modify their own isolate data and data submitted by anybody else that is in the same *user group* as them but not anyone elses. A limited range of *Individual permissions* can be set for each submitter, so their roles can be controlled. A submitter with no specific permissions set has no more power than a standard user.
- Curator can modify data but does not have full control of the database. *Individual permissions* can be set for each curator, so their roles can be controlled. A curator with no specific permissions set has no more power than a standard user.
- Admin has full control of the database, including setting permissions for curators and setting user passwords if built-in authentication is in use.

## 5.2 User groups

User groups allow submitter accounts to be grouped such that the submitter can edit isolates where the sender is either themselves or any member of a user group to which they belong.

## 5.3 Curator permissions

Individual permissions can be set for each curator:

- disable\_access if set to true, this user is completely barred from access.
- modify\_users allowed to add or modify user records. They can change the status of users, but can not revoke admin priveleges from an account. They can also not raise the status of a user to admin level.
- modify\_usergroups allowed to add or modify user groups and add users to these groups.

- set\_user\_passwords allowed to modify other users' passwords (if built-in authentication is in use).
- modify\_loci allowed to add or modify loci.
- modify\_schemes allowed to add or modify schemes.
- modify\_sequences allowed to add sequences to the sequence bin (for isolate databases) or new allele definitions (for sequence definition databases).
- modify\_experiments define new experiments that can be used to group contigs uploaded to the sequence bin.
- modify\_isolates allowed to add or modify isolate records.
- modify\_projects allowed to create projects, modify their descriptions and add or remove isolate records to these.
- modify\_composites allowed to add or modify composite fields (fields made up of other fields, including scheme fields defined in external databases). Composite fields involve defining regular expressions that are evaluated by Perl this can be dangerous so this permission should be granted with discretion.
- modify\_field\_attributes allow user to create or modify secondary field attributes (lookup tables) for isolate record fields.
- modify\_value\_attributes allow user to add or modify secondary field values for isolate record fields.
- modify\_probes allow user to define PCR or hybridization reactions to filter tag scanning.
- tag\_sequences allowed to tag sequences with locus information.
- designate\_alleles allowed to manually designate allele numbers for isolate records.
- modify\_profiles allowed to add or modify scheme profiles (only used in a sequence definitions database).

Permissions can be set by clicking the '?' button next to 'curator permissions' on the curator's interface:

Database: Sp Users: Add   Users: Add   Users: Add   O	Query/u	pdate		)   Curator's page (database)
.ogged in: Keith Jolley (keith). ⊯Log ou	it   Chang	ge password		
Database curator's	inte	rface	- Neisseria	PubMLST
Add, update or delet	te reco	ords		
Record type	Add	Batch Add	Update or de	elete Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
isolates	+	++	query   browse batch upda	
isolate field extended attribute values	+	++	?	Add values for additional isolate field attributes.
projects	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
isolate aliases	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags		scan	?	Tag regions of sequences within the sequence bin with locus information.
Database configurati Table	Add B	atch Add	Update or delete	Comments
loci	+		2	
locus elieses		ankscan	2 Ad	d alternative annual for last. These are also be activitied uso batch add last
locus aliases	+	++		d alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes				ine nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
isolate field extended attributes	+	++	? Def	ine additional attributes to associate with values of a particular isolate record field.

Choose one or more curators from the list (hold down Ctrl to select multiple values). click 'Select'.



Click the appropriate checkboxes to modify permissions. There are also 'All/None' buttons to facilitate quicker selection of options. Click 'Update'.

	og out   Change pass	word					Help 🖉
et curator perm	issions						
Select curator(s)	Action						
Bray, James Christodoulides, Myron	Select						
Christodoulides, Myron Clemence, Marianne							
Diallo, Kanny Exley, Rachel							
Feavers, Ian							
Goodyer-Sait, Lily Halkilahti, Jani							
Haikilanti, Jani	J						
All None							
Check the boxes for the requi	red permissions. (	Jsers with a sta	itus of 'subr	nitter' have a restrict	ted list of allowed	permissions that can be s	selected.
Update permissions				Action			
Permission	Curato	)r	All/None	Update			
CI	emence, Marianne	e Diallo, Kanny	All/None	opullo			
CI modify users		e   Diallo, Kanny		opadio			
Cl modify users modify isolates		e   Diallo, Kanny   		opula			
Cl modify users modify isolates modify projects		e Diallo, Kanny		Cpulle			
Cl modify users modify isolates modify projects modify sequences		e Diallo, Kanny		Cpulle			
Cl modify users modify isolates modify projects modify sequences tag sequences		e Diallo, Kanny V V V					
Cl modify users modify isolates modify projects modify sequences tag sequences designate alleles		e Diallo, Kanny V V V V					
Cl modify users modify isolates modify projects modify sequences tag sequences designate alleles modify usergroups		e Diallo, Kanny V V V V V V V V					
Cl modify users modify isolates modify projects modify sequences tag sequences designate alleles modify usergroups set user passwords		e Diallo, Kanny V V V V V V V					
modify users     Cl       modify isolates     modify isolates       modify projects     modify sequences       tag sequences     designate alleles       modify usergroups     set user passwords       modify loci     modify loci		E Diallo, Kanny					
modify users       Cl         modify isolates       modify isolates         modify projects       modify sequences         tag sequences       designate alleles         modify usergroups       set user passwords         modify loci       modify schemes		Diallo, Kanny           Image: Constraint of the second se					
modify users       Cl         modify isolates       modify isolates         modify projects       modify sequences         tag sequences       designate alleles         modify usergroups       set user passwords         modify loci       modify schemes         modify composites       modify composites		E Diallo, Kanny					
modify users       Cl         modify isolates       modify isolates         modify projects       modify sequences         tag sequences       designate alleles         designate alleles       modify usergroups         set user passwords       modify loci         modify schemes       modify composites         modify field attributes       modify field attributes		Diallo, Kanny Diallo, Kanny V </th <th></th> <th></th> <th></th> <th></th> <th></th>					
modify users       Cl         modify isolates       modify isolates         modify projects       modify sequences         tag sequences       designate alleles         modify usergroups       set user passwords         modify loci       modify schemes         modify composites       modify composites		Diallo, Kanny           Image:					
modify users       Cl         modify isolates       modify isolates         modify projects       modify sequences         tag sequences       designate alleles         designate alleles       modify usergroups         set user passwords       modify loci         modify schemes       modify composites         modify field attributes       modify value attributes		Diallo, Kanny           Diallo, Kanny					
modify users       Cl         modify isolates       modify isolates         modify projects       modify sequences         tag sequences       designate alleles         designate alleles       modify usergroups         set user passwords       modify loci         modify schemes       modify composites         modify field attributes       modify value attributes         modify robes       modify probes		Diallo, Kanny           Diallo, Kanny					
modify users         modify isolates         modify projects         modify sequences         tag sequences         designate alleles         modify usergroups         set user passwords         modify schemes         modify composites         modify reductributes         modify value attributes         modify value attributes		Diallo, Kanny           Diallo, Kanny           Image: Comparison of the state of the					

The 'disable access' option provides a quick way to disable access to a curator. This will not be selected by the 'All/None' buttons.

# 5.4 Locus and scheme permissions (sequence definition database)

To be allowed to define alleles or scheme profiles, curators must be granted specific permission for each locus and scheme by adding their user id number to the 'locus curator' and 'scheme curator' lists.

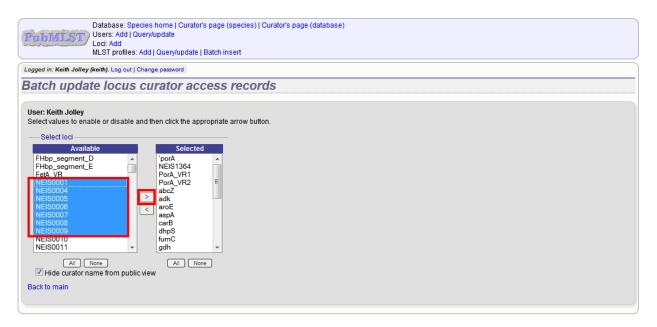
The easiest way to modify these lists is to use the batch update link next to 'locus curator control list' and 'scheme curator control list':

Users: Add   Query/up Loci: Add MLST profiles: Add   Q		update   B	atch insert	
ogged in: Keith Jolley (keith). 🕒 Log out   Change	e passv	vord		
atabase curator's inter	fac	:e - N	eisseria locus/s	sequence definitions
📝 Add, update or delete reco	rds			
Record type	Add	Batch Add	Update or delete	Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
locus descriptions	+	++	?	
scheme curator control list	+	++	query batch	Define which curators can add or update profiles for particular schemes.
locus curator control list	+	++	query batch	Define which curators can add or update sequences for particular loci.
sequences (all loci)	+	++   FASTA	?	
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
PubMed links (to sequences)	+	++	?	
Bexsero Antigen Sequence Typing (BAST) profiles	+	++	query   browse   list   batch update	
MLST profiles	+	++	query   browse   list   batch update	
rplF species profiles	+	++	query   browse   list   batch update	

#### Select the curator from the list:

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loc: Add MLST profiles: Add   Query/update   Batch insert
Logged in: Keith Jolley (keith). Log out   Change password
Batch update locus curator access records
Select user
The user status must also be set to curator for permissions to work.
USers: Jolley, Keith (keith)

Then select loci/schemes that the user is allowed to curate in the left hand 'Available' list, and click the right button to move these to the 'Selected' list:



If you uncheck the 'Hide curator name from public view' checkbox, the curator name and E-mail address will appear alongside loci in the download table on the website.

# 5.5 Controlling access

### 5.5.1 Restricting particular configurations to specific user accounts

Suppose you only wanted specific users to access a database configuration.

In the config.xml, add the following directive:

default\_access="deny"

This tells BIGSdb to deny access to anybody unless their account name appears within a file called users.allow within the config directory. The users.allow file should contain one username per line.

Alternatively, you can deny access to specific users, while allowing every other authenticated user. In config.xml, add the following directive:

default\_access="allow"

This tells BIGSdb to allow access to anybody unless their account name appears within a file called users.deny within the config directory. The users.deny file should contain one username per line.

## 5.6 Setting user passwords

Please note that these instructions only apply if using the built-in BIGSdb authentication system.

If you are an administrator or a curator with specific permission to change other users' passwords, you should see a link to 'set user passwords' at the bottom of the curator's index page. Click this.

Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
		tabank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
cus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlinks and searches of these databases when you query sequences or profiles in this database.
client database loci	+	++	?	Define loci that are used in client databases.
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
cheme group scheme members	+	++	?	Defines which schemes belong to a group.
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.

Select the appropriate user from the drop-down list box and enter the new password twice where prompted.

PubMLST / Users: Add   Qu	cies home   Curator's page (species)   Curator's page (database) Jery/update Query/update   Batch insert
Logged in: Keith Jolley (keith). Log out   C	hange password
Set user password	
User: Jolley, Keith (kei New password: •••••• Retype password: ••••••	th)
Relype password.	Set password

Click 'Set password' and the password will be updated.

# 5.7 Setting the first user password

To set the first administrator's password for a new database, use the add\_user.pl script found in the scripts/maintenance directory:

add\_user.pl [-a] -d <dbase> -n <username> -p <password>

The first user account needs to be added to the database *manually*.

# 5.8 Enabling plugins

Some plugins can be enabled/disabled for specific databases. If you look in the get\_attributes function of the specific plugin file and see a value for system\_flag, this value can be used in the system tag of the database configuration XML file to enable the plugin.

For example, the get\_attributes function of the BURST plugin looks like:

```
sub get_attributes {
      my %att = (
                        => 'BURST',
              name
              author => 'Keith Jolley',
              affiliation => 'University of Oxford, UK',
              email => 'keith.jolley@zoo.ox.ac.uk',
              description => 'Perform BURST cluster analysis on query results query results',
              category => 'Cluster',
              buttontext => 'BURST',
              menutext => 'BURST',
              module
                         => 'BURST',
              version
                        => '1.0.0',
              dbtype => 'isolates, sequences',
              section => 'postquery',
order => 10,
              system_flag => 'BURST',
                       => 'query',
              input
              requires => 'mogrify',
                        => 2,
              min
                         => 1000
              max
      );
      return \%att;
```

The 'system\_flag' attribute is set to 'BURST', so this plugin can be enabled for a database by adding:

BURST="yes"

to the system tag of the database XML file. If the system\_flag value is not defined then the plugin is always enabled if it is installed on the system.

# 5.9 Temporarily disabling database updates

There may be instances where it is necessary to temporarily disable database updates. This may be during periods of server or database maintenance, for instance when running on a backup database server.

Updates can be disabled on a global or database-specific level.

### 5.9.1 Global

In the /etc/bigsdb/bigsdb.conf file, add the following line:

disable\_updates=yes

An optional message can also be displayed by adding a disable\_update\_message value, e.g.

disable\_update\_message=The server is currently undergoing maintenance.

## 5.9.2 Database-specific

The same attributes described above for use in the bigsdb.conf file can also be used within the system tag of the database config.xml file, e.g.

```
<system
db="bigsdb_neisseria"
dbtype="isolates"
...
disable_updates="yes"
disable_update_message="The server is currently undergoing maintenance."
```

# 5.10 Host mapping

During periods of server maintenance, it may be necessary to map a database host to an alternative server. This would allow a backup database server to be used while the primary database server is unavailable. In this scenario, you would probably also want to *disable updates*.

Host mapping can be achieved by editing the /etc/bigsdb/host\_mapping.conf file. Each host mapping is placed on a single line, with the current server followed by any amount of whitespace and then the new mapped host, e.g.

#Existing_host	Mapped_host
server1	server2
localhost	server2

[Lines beginning with a hash are comments and are ignored.]

This configuration would use server2 instead of server 1 or localhost wherever they are defined in the database configuration (either host attribute in the database config.xml file, or within the loci or schemes tables).

## 5.11 Improving performance

### 5.11.1 Use mod\_perl

The single biggest improvement to speed can be obtained by running BIGSdb under mod\_perl. There's very little point trying anything else until you have mod\_perl set up and running - this can improve start-up performance a hundred-fold since the script isn't compiled on each page access but persists in memory.

## 5.11.2 Cache scheme definitions within an isolate database

If you have a large number of allelic profiles defined for a scheme, you can cache these definitions within an isolate database to speed up querying of isolates by scheme criteria (e.g. by ST for a MLST scheme).

To do this use the update\_scheme\_caches.pl script found in the scripts/maintenance directory, e.g. to cache all schemes in the pubmlst\_bigsdb\_neisseria\_isolates database

update\_scheme\_caches.pl -d pubmlst\_bigsdb\_neisseria\_isolates

This script creates indexed tables within the isolate database called temp\_scheme\_X and temp\_isolates\_scheme\_fields\_1 (where X is the scheme\_id). If these table aren't present, they are created as temporary tables every time a query is performed that requires a join against scheme definition data. This requires importing all profile definitions from the definitions database and determining scheme field values for all isolates. This may sound like it would be slow but caching only has a noticeable effect once you have >5000 profiles.

Note that you will need to run this script periodically as a CRON job to refresh the cache. Admins can also refresh the caches manually from a link on the curators' page. This link is only present if the caches have been previously generated.

allele designations				
		++		? Allele designations can be set within the isolate table functions.
sequences		++		? The sequence bin holds sequence contigs from any source.
accession number links	+	++		? Associate sequences with Genbank/EMBL accession number.
experiments	+	++		? Set up experiments to which sequences in the bin can belong.
experiment sequences				? Add links associating sequences to experiments.
sequence tags		scan		? Tag regions of sequences within the sequence bin with locus information.
Submissions				
how closed submissions				
Database configuratio				
Table	Add∣P	atch Add I	Jpdate or delete	el Comments
	_			
loci	+	++	?	
loci	+ datab	++ ank scan	?	
loci locus aliases	+ datab +	++ ank scan ++		Add alternative names for loci. These can also be set when you batch add loci.
loci locus aliases PCR reactions	+ datab + +	++ ank scan ++ ++	?	Add alternative names for loci. These can also be set when you batch add loci. Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
loci locus aliases PCR reactions nucleotide probes	+ datab + + +	++ ank scan ++ ++ ++	? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci.
loci PCR reactions nucleotide probes solate field extended attributes	+ datab + + +	++ ank scan ++ ++	? ? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. Define additional attributes to associate with values of a particular isolate record field.
loci PCR reactions nucleotide probes solate field extended attributes composite fields	+ datab + + + + + +	++ ank scan ++ ++ ++ ++	? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. Define additional attributes to associate with values of a particular isolate record field. Used to construct composite fields consisting of fields from isolate. Icoid or scheme fields.
loci locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes	+ datab + + + + + + + + +	+++ ank scan +++ ++ ++ ++ ++ ++ ++	? ? ? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. Define additional attributes to associate with values of a particular isolate record field. Used to construct composite fields consisting of fields from isolate, loci or scheme fields. Define attributes that can be set for contigs in the sequence bin.
loci locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes schemes	+ datab + + + + + + + + + + +	++ ank scan ++ ++ ++ ++ ++ ++ ++ ++ ++ +	? ? ? ? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up in <i>silloo</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define additional attributes to associate with values of a particular isolate record field. Used to construct composite fields consisting of fields from isolate, loci or scheme fields. Define additional attributes to be set for contigs in the sequence bin. Define additional generation of loci, e.g. MLST.
loci locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes schemes scheme members	+ datab + + + + + + + + + + +	++	? ? ? ? ? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci. Define additional attributes to associate with values of a particular isolate record field. Used to construct composite fields consisting of fields from isolate, loci or scheme fields. Define attributes that can be set for contigs in the sequence bin. Describes schemes consisting of collections of loci, e.g. MLST. Defines which loci belong to a scheme.
loci locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes schemes scheme members scheme fields	+ datab + + + + + + + + + + + + + +	+++ ank scan 444 444 444 444 444 444 444 444 444 4	? ? ? ? ? ? ? ? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. Define additional attributes to associate with values of a particular isolate record field. Used to construct composite fields consisting of fields from isolate. Joci or scheme fields. Define attributes that can be set for contigs in the sequence bin. Describes schemes consisting of collections of loci, e.g. MLST. Defines which fields belong to a scheme.
loci locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes schemes scheme members scheme groups	+ datab + + + + + + + + + + + + + + +	+++ ank scan +++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ +	? ? ? ? ? ? ? ? ? ? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up in <i>silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. Define additional attributes to associate with values of a particular isolate record field. Used to construct composite fields consisting of fields from isolate, loci or scheme fields. Define additional stributes that can be set for contigs in the sequence bin. Describes schemes consisting of collections of loci, e.g. MLST. Defines which fields belong to a scheme. Defines which fields belong to a scheme. Describes groups in to which schemes can belong - groups can also belong to other groups.
loci locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes schemes scheme members scheme fields	+ datab + + + + + + + + + + + + + + + + +	+++ ank scan 444 444 444 444 444 444 444 444 444 4	? ? ? ? ? ? ? ? ?	Add alternative names for loci. These can also be set when you batch add loci. Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. Define additional attributes to associate with values of a particular isolate record field. Used to construct composite fields consisting of fields from isolate. Joci or scheme fields. Define attributes that can be set for contigs in the sequence bin. Describes schemes consisting of collections of loci, e.g. MLST. Defines which fields belong to a scheme.

You can also set cache\_schemes="yes" in the system tag of config.xml to enable automatic refreshing of the caches when batch adding new isolates (you should still periodically run the update\_scheme\_caches.pl script via CRON to ensure any changes in the sequence definition database are picked up).

If queries are taking longer than 5 seconds to perform and a cache is not in place, you will see a warning message in bigsdb.log suggesting that the caches be set up. Unless you see this warning regularly, you probably don't need to do this.

### 5.11.3 Use materialized views for scheme definitions

Because of the way BIGSdb allows any number of profile schemes to be set up, the data are stored in a normalised manner in multiple tables. A database view, e.g. scheme\_1, is created that joins these tables so that they can be queried as you would a single table. A view, however, is only a pre-selected query rather than a physical table and you can not index columns on it to optimise query performance.

A materialized view is a real table that is created from the view and refreshed every time the data in the underlying view changes. Because it is a real table, the database doesn't need to perform these joins every time it is queried and indexes can be set up on it, both of which greatly speeds up querying.

To use materialized views within a sequef database set the following attribute in the system tag of the XML description file:

```
materialized_views="yes"
```

You will then need to run the 'configuration repair' function at the bottom of the administrator's main curation page for each scheme. This rebuilds the view and creates a materialized view called mv\_scheme\_X. This materialized view is updated automatically whenever profile data are added or altered via the web interface.

If you want an isolate database to benefit from this materialized view, make sure you put 'mv\_scheme\_X' (where X is the scheme id) in the dbase\_table field (rather than 'scheme\_X') when setting up the scheme in the isolate database configuration.

Please note that if you make changes to your profile data by means other than the web interface then the materialized view will not be updated. You can update it by running the following SQL command:

SELECT refresh\_matview('mv\_scheme\_X');

The materialized view is used, for example, for looking up a ST from a profile and vice-versa. Significant speed improvements will only be realised if you have lots of profiles (>5000) and you are doing lots of lookups, e.g. displaying more than the default 25 records per page.

## 5.11.4 Use a ramdisk for the secure temporary directory

If you are running BIGSdb on a large server with lots of RAM, you could use some of this as a ramdisk for temporary files. Debian/Ubuntu systems make available up to half the system RAM as a ramdisk mounted under /run/shm (or /dev/shm) by default. Set the secure\_tmp\_dir to this RAM disk and you should see significant improvement in operations requiring the writing of lots of temporary files, e.g. tag scanning and the Genome Comparator plugin. This is only likely to be appropriate if you have very large amounts of RAM available. As an example, the server hosting the PubMLST databases is a dedicated machine with 1TB RAM with temporary files rarely using more than 50GB space.

# 5.12 Dataset partitioning

## 5.12.1 Sets

Sets provide a means to partition the database in to manageable units that can appear as smaller databases to an end user. Sets can include constrained groups of isolates, loci, and schemes from the complete database and also include additional metadata fields only applicable to that set.

#### See also:

Sets (concept)

## 5.12.2 Configuration of sets

First sets need to be enabled in the XML configuration file (config.xml) of the database. Add the following attribute to the system tag:

sets="yes"

With this attribute, the curation interface now has options to add sets, and then add loci or schemes to these sets.

ancie ucorgitationo				Allete designations can be set within the isolate table fonctions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags		scan	?	Tag regions of sequences within the sequence bin with locus information.
Database configuration	on			
Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
	data	abank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci
solate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.
sequence attributes	+	++	?	Define attributes that can be set for contigs in the sequence bin.
sets	+	++	?	Sets describe a collection of loci and schemes that can be treated like a stand-alone database.
set loci	+	++	?	Add loci to sets.
set schemes	+	++	?	Add schemes to sets.
set metadata	+	++	?	Add metadata collection to sets.
setview	+	++	?	Set database views linked to sets.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
cheme group scheme members	+	++	?	Defines which schemes belong to a group.
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.
				log on or change an existing password. d schemes and that required helper applications are properly installed.

The name of a locus or scheme to use within a set can be defined in the set\_name field when adding loci or schemes to a set. Common names can also be set for loci - equivalent to the common name used within the loci table.

Now when a user goes to the contents page of the database they will be presented with a dropdown menu of datasets and can choose either the 'whole database' or a specific set. This selection is remembered between sessions.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	Database submissions
Neisseria PubMLST database	
	ent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence late may be submitted to this database and consequently it should be noted that it does not
Datasets This database contains multiple datasets. You can choose to display a single set or th Please select: Whole database   Choose	e whole database.
Query database • Search database - advanced queries. • Browse database - peruse all records. • Search by combinations of loci (profiles) - including partial matching. • List query - find isolates by matching a field to an entered list.	Option settings   Set general options - including isolate table field handling.  Set display and query options for locus, schemes or scheme fields.  General information  Isolates: 20902 Last updated: 2015-06-30 Update history About BIGSdb
Breakdown     Single field     Two field     Unique combinations     Scheme and alleles     Publications     Sequence bin     Sequence bin	Analysis          • Codon usage         • Presence/absence status of loci         • Genome comparator         • BLAST           • Miscellaneous         • Description of database fields         • Description of database fields

Alternatively, a specific set can be selected within the XML config file so that only a specific set is available when accessed via that configuration. In that case, the user would be unaware that the database contains anything other than the loci and schemes available within the set.

To specify this, add the following attributute to the system tag:

set\_id="1"

where the value is the name of the set.

**Note:** If the set\_id attribute is set, database configuration settings in the curator's interface are disabled. This is because when the configuration is constrained to a set, only loci and schemes already added to the set are visible, so functionality to edit schemes/loci would become very confusing. To modify these settings, you either need to access the interface from a different configuration, i.e. an alternative config.xml with the set\_id attribute not set, or temporarily remove the set\_id directive from the current config.xml while making configuration changes.

### 5.12.3 Set metadata

Additional metadata fields can be set within the XML configuration file. They are specified as belonging to a metaset by prefixing the field name with 'meta\_NAME:' where NAME is the name of the metaset, e.g.

```
<field type="text" required="no" length="30" maindisplay="no"
optlist="yes">meta_1:clinical_outcome
<optlist>
<option>no sequeleae</option>
```

```
<option>hearing loss</option>
  <option>amputation</option>
  <option>death</option>
  </optlist>
</field>
```

Metaset fields can be defined just like any other *provenance field* and their position in the output is determined by their position in the XML file.

Metaset fields can then be added to a set using the 'Add set metadata' link on the curator's page.

isolates		++	au and has		
	1			owse   list   update	
isolate field extended attribute values	+	++		?	Add values for additional isolate field attributes.
projects	+	++		?	Set up projects to which isolates can belong.
project members	+	++		?	Add isolates to projects.
isolate aliases	+	++		?	Add alternative names for isolates.
PubMed links	+	++		?	
allele designations		++		?	Allele designations can be set within the isolate table functions.
sequences		++		?	The sequence bin holds sequence contigs from any source.
accession number links	+	++		?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++		?	Set up experiments to which sequences in the bin can belong.
experiment sequences				?	Add links associating sequences to experiments.
sequence tags		scan		?	Tag regions of sequences within the sequence bin with locus information.
Table	Auu		Update or delete		Comments
			-		
loci	+	++	?		
		++ bank scan			
locus aliases		bank scan ++	?		names for loci. These can also be set when you batch add loci.
locus aliases PCR reactions	data + +	bank scan ++ ++		Set up in silico	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
locus aliases PCR reactions nucleotide probes	data + + +	bank scan ++ ++ ++	?	Set up <i>in silico</i> Define nucleotic	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci.
locus aliases PCR reactions nucleotide probes solate field extended attributes	data + + + +	bank scan ++ ++	?	Set up in silico I Define nucleotic Define additions	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields	data + + + + +	bank scan ++ ++ ++ ++	? ? ? ?	Set up in silico l Define nucleotic Define additions Used to construct	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci, de probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci, al attributes to associate with values of a particular isolate record field, ct composite fields consisting of fields from isolate, loci or scheme fields.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes	data + + + + + +	bank scan ++ ++ ++ ++ ++	? ? ? ? ?	Set up in silico I Define nucleotic Define additions Used to construct Define attributes	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. ct composite fields consisting of fields from isolate, loci or scheme fields. s that can be set for contigs in the sequence bin.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes sets	data + + + + + + + + +	bank scan ++ ++ ++ ++ ++ ++ ++	? ? ? ? ?	Set up in silico l Define nucleotic Define additiona Used to construc Define attributes Sets describe a	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. at composite fields consisting of fields from isolate, loci or scheme fields. I shat can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-slone database.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes sets sets set loci	data + + + + + + + + + + +	bank scan +++ +++ +++ +++ +++ +++ +++	? ? ? ? ? ? ?	Set up in silico l Define nucleotio Define additions Used to construct Define attributes Sets describe a Add loci to sets.	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in ailico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. at composite fields consisting of fields from isolate, loci or scheme fields. s that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database.
locus aliases PCR reactions nucleotide probes tolate field extended attributes composite fields sequence attributes sets set loci set loci set schemes	data + + + + + + + + + + + + +	bank scan +++ +++ +++ +++ +++ +++ +++ +++ +++	? ? ? ? ? ? ? ? ? ?	Set up in silico l Define nucleotio Define additions Used to construc Define attributes Sets describe a Add loci to sets. Add schemes to	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. d composite fields consisting of fields from isolate, loci or scheme fields. s that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database. sets.
locus aliases PCR reactions nucleotide probes olate field extended attributes composite fields sequence attributes sets set loci set schemes set metadata	data + + + + + + + + + + +	bank scan ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++	? ? ? ? ? ? ? ? ? ? ? ? ?	Set up in silico l Define nucleotic Define additions Used to construc Define attributes Sets describe a Add loci to sets. Add schemes to Add metadata o	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silloo</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. t composite fields consisting of fields from isolate, loci or scheme fields. s that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database. sets.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes sets set loci set schemes set metadata set view	data + + + + + + + + + + + + + + +	bank scan ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++	? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	Set up in silico I Define nucleotic Define additiona Used to construc Define attributes Sets describe a Add loci to sets. Add schemes to Add metadata o Set database vie	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. d composite fields consisting of fields from isolate, loci or scheme fields. Is that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database. sets. sets. sets. we linked to sets.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes sets sets set loci set schemes set metadata set view schemes	data + + + + + + + + + + + + + + + +	bank scan ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	Set up in silico I Define nucleotic Define additions Used to construc Define attributes Sets describe a Add loci to sets. Add schemes to Add metadata o Set database vie Describes schem	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in ailico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. at composite fields consisting of fields from isolate, loci or scheme fields. Is that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database. sets. collection to sets. eves linked to sets. hes consisting of collections of loci, e.g. MLST.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes sets sets set loci set schemes set metadata set view schemes scheme members	data + + + + + + + + + + + + + + + + + +	bank scan ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	Set up in silico I Define nucleotic Define additions Used to construct Sets describe a Add loci to sets. Add schemes to Add metadata o Set database vic Describes schem Defines which lo	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in ailico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. at composite fields consisting of fields from isolate, loci or scheme fields. s that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database. sets. collection to sets. ews linked to sets. ews linked to sets. beiong of oclections of loci, e.g. MLST. bol belong to a scheme.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes set loci set schemes set schemes set wew schemes scheme members scheme fields	data + + + + + + + + + + + + + + + + + +	bank scan ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	Set up in silico I Define nucleotic Define additionu Used to construc Define attributes Sets desoribe a Add loci to sets. Add schemes to Add metadata o Set database vir Desoribes schem Defines which fü	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silloo</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. t composite fields consisting of fields from isolate, loci or scheme fields. s that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database. sets. collection to sets. wers linked to sets. nes consisting of collections of loci, e.g. MLST. col belong to a scheme. elds belong to a scheme.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes sets set loci set schemes set metadata set view schemes scheme fields scheme groups	data + + + + + + + + + + + + + + + + + + +	bank scan ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	Set up in silico I Define nucleotic Define additionu Used to construc Define attributes Sets describe a Add loci to sets. Add schemes to Add schemes to Add metadata o Set database vito Describes schem Defines which fü Describes group	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. t composite fields consisting of fields from isolate, loci or scheme fields. t that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database. sets. collection to sets. eves linked to sets. eves linked to sets. eves of collections of loci, e.g. MLST. col belong to a scheme. elds belong to a scheme. s in to which schemes can belong - groups can also belong to other groups.
locus aliases PCR reactions nucleotide probes solate field extended attributes composite fields sequence attributes set loci set schemes set netadata set view schemes scheme members scheme fields	data + + + + + + + + + + + + + + + + + + +	bank scan ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++	? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?	Set up in silico I Define nucleotic Define additiona Used to construct Sets describe a Add loci to sets. Add schemes to Set database vir Describes schem Defines which für Describes group Defines which so	PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. de probes for <i>in silloo</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. al attributes to associate with values of a particular isolate record field. t composite fields consisting of fields from isolate, loci or scheme fields. s that can be set for contigs in the sequence bin. collection of loci and schemes that can be treated like a stand-alone database. sets. collection to sets. wers linked to sets. nes consisting of collections of loci, e.g. MLST. col belong to a scheme. elds belong to a scheme.

A new database table needs to be added for each metaset. This should contain all the fields defined for a metaset. The table should also contain an isolate\_id field to act as the foreign key linking to the isolate table, e.g. the SQL would look something like the following:

```
CREATE TABLE meta_1 (
isolate_id integer NOT NULL,
town text,
clinical_outcome text,
PRIMARY KEY (isolate_id),
CONSTRAINT m1_isolate_id FOREIGN KEY (isolate_id) REFERENCES isolates
ON DELETE CASCADE
ON UPDATE CASCADE
);
GRANT SELECT,UPDATE,INSERT,DELETE ON meta_1 TO apache;
```

The above creates the database table for a metaset called '1', defining new text fields for 'town' and 'clinical\_outcome'.

## 5.12.4 Set views

Finally the isolate record table can be partitoned using database views and these views associated with a set. Create views using something like the following:

```
CREATE VIEW spneumoniae AS SELECT * FROM isolates WHERE species = 'Streptococcus pneumoniae';
GRANT SELECT ON spneumoniae TO apache;
```

Add the available views to the XML file as a comma separated list in the system tag 'views' attribute:

```
<system
.....
sets="yes"
views="spneumoniae,saureus"
>
</system>
```

Set the view to the set by using the 'Add set view' link on the curator's page.

## 5.12.5 Using only defined sets

The only\_sets attribute can be set to 'yes' to disable the option for 'Whole database' so that only sets can be viewed, e.g.

```
<system
.....
sets="yes"
only_sets="yes"
>
</system>
```

# 5.13 Adding new loci

#### See also:

Loci (concept)

### 5.13.1 Sequence definition databases

#### Single locus

Click the add (+) loci link on the curator's interface contents page.

Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
		itabank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
ocus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlinks and searches of thes databases when you query sequences or profiles in this database.
client database loci	+	++	?	Define loci that are used in client databases.
lient database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
scheme group scheme members	+	++	?	Defines which schemes belong to a group.
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.

Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

- id The name of the locus.
  - Allowed: any value starting with a letter or underscore.
- data\_type Describes whether the locus is defined by nucleotide or peptide sequence.
  - Allowed: DNA/peptide.
- allele\_id\_format The format for allele identifiers.
  - Allowed: integer/text.
- length\_varies Sets whether alleles can vary in length.
  - Allowed: true/false.
- coding\_sequence Sets whether the locus codes for a protein.
  - Allowed: true/false.
- formatted\_name Name with HTML formatting (optional).
  - This allows you to add formatting such as bold, italic, underline and superscripting to locus names as they
    appear in the web interface.
  - Allowed: valid HTML.
- common\_name The common name for the locus (optional).
  - Allowed: any value.
- formatted\_common\_name Common name with HTML formatting (optional).
  - Allowed: valid HTML.
- allele\_id\_regex Regular expression to enforce allele id naming (optional).
  - ^: the beginning of the string
  - \$:the end of the string
  - d: digit

- D: non-digit
- s: white space character
- S: non white space character
- w: alpha-numeric plus '\_'
- .: any character
- \*: 0 or more of previous character
- +: 1 or more of previous character
- e.g. ^Fd-d+\$ states that an allele name must begin with a F followed by a single digit, then a dash, then one or more digits, e.g. F1-12
- length Standard length of locus (required if length\_varies is set to false.
  - Allowed: any integer.
- min\_length Minimum length of locus (optional).
  - Allowed: any integer.
- max\_length Maximum length of locus (optional).
  - Allowed: any integer (larger than the minimum length).
- orf Open reading frame of locus (optional).
  - 1-3 are the forward reading frame, 4-6 are the reverse reading frames.
  - Allowed: 1-6.
- genome\_position The start position of the locus on a reference genome (optional).
  - Allowed: any integer.
- match\_longest Specifies whether in a sequence query to only return the longest match (optional).
  - This is useful for some loci that can have some sequences shorter than others, e.g. peptide loci defining antigenic loops. This can lead to instances of one sequence being longer than another but otherwise being identical. In these cases, usually the longer sequence is the one that should be matched.
  - Allowed: true/false.
- full\_name Full name of the locus (optional).
  - Allowed: any value.
- product Name of gene product (optional).
  - Allowed: Any value.
- description Description of the locus (optional).
  - Allowed: any value.
- aliases Alternative names for the locus (optional).
  - Enter each alias on a separate line in the text box.
  - Allowed: any value.
- pubmed\_ids PubMed ids of publications describing the locus (optional).
  - Enter each PubMed id on a separate line in the text box.
  - Allowed: any integer.

- links Hyperlinks pointing to additional resources to display in the locus description (optional).
  - Enter each link on a separate line in the format with the URL first, followed by a | then the description (URL/description).

#### **Batch adding**

Click the batch add (++) loci link on the curator's interface contents page.

Table	Add	Batch Add	Update or delete	Comments			
loci	+	++	?				
	da	itabank					
		scan					
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.			
locus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.			
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlinks and searches of these			
				databases when you query sequences or profiles in this database.			
client database loci	+	++	?	Define loci that are used in client databases.			
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.			
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.			
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.			
scheme members	+	++	?	Defines which loci belong to a scheme.			
scheme fields	+	++	?	Defines which fields belong to a scheme.			
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.			
scheme group scheme members	+	++	?	Defines which schemes belong to a group.			
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.			
	k - Ch	ecks data	base connecti	able them to log on or change an existing password. ity for loci and schemes and that required helper applications are properly installed.			

#### Click the link to download a header line for an Excel spreadsheet:

1		
Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add		
MLST profiles: Add   Query/update   Batch insert		
Logged in: Keith Jolley (keith). Log out   Change password		Toggle: i
Batch insert loci		
This page allows you to upload locus data as tab-delimited text or copied from a spreadsheet.  • Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.  • Download tab-delimited header for your spreadsheet - use Paste special → text to paste the data.  • Download submission template (xlsx format)		
Paste in tab-delimited text (include a field header line).	Action	
r aste in tab delinine ditext (include a neta neader line).		
	Reset Submit	
h.		
Back		

Fill in the spreadsheet using the fields described for *adding single loci*.

Fill in the spreadsheet fields using the table above as a guide, then paste the completed table into the web form and press 'Submit query'.

## 5.13.2 Isolate databases

### Single locus

Click the add (+) loci link on the curator's interface contents page.

Users: Add   C Isolates: Add			atch insert		
ed in: Keith Jolley (keith). 🗭Log ou	t   Chang	e password			
tabase curator's	inte	rface -	- Neisser	ia PubN	<i>ILST</i>
Add, update or delet	e reco	ords			
Record type	Add	Batch Add	Update o	or delete	Comments
users	+	++		?	
user groups	+	++	1	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++		?	Add users to groups for setting access permissions.
curator permissions			1	?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
isolates	+	++		wse list  update	
solate field extended attribute values	+	++	1	?	Add values for additional isolate field attributes.
projects	+	++	1	?	Set up projects to which isolates can belong.
project members	+	++	1	?	Add isolates to projects.
isolate aliases	+	++	1	?	Add alternative names for isolates.
PubMed links	+	++	1	?	
allele designations		++	1	?	Allele designations can be set within the isolate table functions.
sequences		++	1	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	1	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	1	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			1	?	Add links associating sequences to experiments.
sequence tags		scan	1	?	Tag regions of sequences within the sequence bin with locus information.
Database configurati					
Table			Jpdate or delete		Comments
loci	+	++	?		
logua aliagoa	databa	ank scan	2	Add alternative	exercise lasi. These was also be estudion you batch add lasi
locus aliases	+	++			names for loci. These can also be set when you batch add loci.
PCR reactions	+	++			PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes olate field extended attributes	+	++	2		de probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
composite fields		++	?		al attributes to associate with values of a particular isolate record field.
	+		1	Used to constru	ct composite fields consisting of fields from isolate, loci or scheme fields.

Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

PubMLST Users: Add	pecies home   Curator's page (species)   Cura Query/update s: Add   Query/update   Batch insert	itor's page (database)	
Logged in: Keith Jolley (keith). 🕒 Log o	ut   Change password		Toggle: 🚯
Add new locus			 
Please fill in the fields below - re	quired fields are marked with an exclamation n	nark (I)	
- Record			Action
id:	2111		Reset Submit
data type:			
allele id format:			
	○ true ● false 1		
coding sequence:			
	Keith Jolley (keith)		
date entered: datestamp:			
formatted name:	2015-07-01	0	
common name:		0	
formatted common name:		0	
allele id regex:		ŏ	
length:	0		
min length:			
max length:	÷ 0		
orf:	<b>•</b> 0		
genome position:	÷ 0		
match longest:	🛇 true 🔘 false 🕦		
complete cds:	🛇 true 🔘 false 🟮		
full name:			
product			
description:			
description.			

- id The name of the locus
  - Allowed: any value starting with a letter or underscore.
- data\_type Describes whether the locus is defined by nucleotide or peptide sequence.
  - Allowed: DNA/peptide.
- allele\_id\_format The format for allele identifiers.
  - Allowed: integer/text.
- length\_varies Sets whether alleles can vary in length.
  - Allowed: true/false.
- coding\_sequence Sets whether the locus codes for a protein.
  - Allowed: true/false.
- flag\_table Set to true to specify that the sequence definition database contains an allele flag table (which is the case for BIGSdb version 1.4 onwards).
  - Allowed: true/false.
- isolate\_display Sets how alleles for this locus are displayed in a detailed isolate record this can be overridden by user preference.
  - Allowed: allele only/sequence/hide.
- main\_display Sets whether or not alleles for this locus are displayed in a main results table by default this can be overridden by user preference.

- Allowed: true/false.
- query\_field Sets whether or not alleles for this locus can be used in queries by default this can be overridden by user preference.
  - Allowed: true/false.
- analysis Sets whether or not alleles for this locus can be used in analysis functions by default this can be overridden by user preference.
  - Allowed: true/false.
- formatted\_name Name with HTML formatting (optional).
  - This allows you to add formatting such as bold, italic, underline and superscripting to locus names as they
    appear in the web interface.
  - Allowed: valid HTML.
- common\_name The common name for the locus (optional).
  - Allowed: any value.
- formatted\_common\_name Common name with HTML formatting (optional).
  - Allowed: valid HTML.
- allele\_id\_regex Regular expression to enforce allele id naming.
  - ^: the beginning of the string
  - \$:the end of the string
  - d: digit
  - D: non-digit
  - s: white space character
  - S: non white space character
  - w: alpha-numeric plus '\_'
  - .: any character
  - \*: 0 or more of previous character
  - +: 1 or more of previous character
  - e.g. ^Fd-d+\$ states that an allele name must begin with a F followed by a single digit, then a dash, then one or more digits, e.g. F1-12
- length Standard length of locus (required if length\_varies is set to false).
  - Allowed: any integer.
- orf Open reading frame of locus (optional). 1-3 are the forward reading frame, 4-6 are the reverse reading frames.
  - Allowed: 1-6.
- genome\_position The start position of the locus on a reference genome.
  - Allowed: any integer.
- match\_longest Only select the longest exact match when tagging/querying.

- This is useful for some loci that can have some sequences shorter than others, e.g. peptide loci defining antigenic loops. This can lead to instances of one sequence being longer than another but otherwise being identical. In these cases, usually the longer sequence is the one that should be matched.
- Allowed: true/false.
- reference\_sequence Sequence used by the automated sequence comparison algorithms to identify sequences matching this locus. This is only used if a sequence definition database has not been set up for this locus.
- pcr\_filter Set to true if this locus is further defined by genome filtering using in silico PCR.
  - Allowed: true/false.
- probe\_filter Set to true if this locus is further defined by genome filtering using in silico hybdridization.
  - Allowed: true/false.
- dbase\_name Name of database (system name).
  - Allowed: any text.
- dbase\_host Resolved name of IP address of database host leave blank if running on the same machine as the isolate database.
  - Allowed: network address, e.g. 129.67.26.52 or zoo-oban.zoo.ox.ac.uk
- dbase\_port Network port on which the sequence definition database server is listening leave blank unless using a non-standard port (5432).
  - Allowed: integer.
- dbase\_user Name of user with permission to access the sequence definition database depending on the database configuration you may be able to leave this blank.
  - Allowed: any text (no spaces).
- dbase\_password Password of database user again depending on the database configuration you may be able to leave this blank.
  - Allowed: any text (no spaces).
- dbase\_table Table in the sequence definition database that contains allele sequences for this locus. If the definition database uses BIGSdb this will be 'sequences'.
  - Allowed: any text (no spaces).
- dbase\_id\_field Primary field in sequence database that defines allele. If the definition database uses BIGSdb this will be 'allele\_id'.
  - Allowed: any text (no spaces).
- dbase\_id2\_field Secondary field in sequence database that defines alleles. If dbase\_id\_field uniquely defines alleles for this locus then this should be left blank. If the definition database uses BIGSdb this will be 'locus'.
  - Allowed: any text (no spaces).
- dbase\_id2\_value Secondary field value in sequence database that defines alleles. If dbase\_id\_field uniquely defines alleles for this locus then this should be left blank. If the definition database uses BIGSdb this will be the name of the locus used in the id field
  - Allowed: any text (no spaces).
- dbase\_seq\_field Field in sequence database containing allele sequence. If the definition database uses BIGSdb this will be 'sequence'.
  - Allowed: any text (no spaces).

- description\_url The URL used to hyperlink to locus information in the isolate information page. This can either be a relative (e.g. /cgi-bin/...) or an absolute (containing http://) URL.
  - Allowed: any valid URL.
- url The URL used to hyperlink to information about the allele. This can either be a relative or absolute URL. If [?] (including the square brackets) is included then this will be substituted for the allele value in the resultant URL. To link to the appropriate allele info page on a corresponding seqdef database you would need something like /cgi-bin/bigsdb/bigsdb.pl?db=pubmlst\_neisseria\_seqdef&page=alleleInfo&locus=abcZ&allele\_id=[?].
  - Allowed: any valid URL.
- submission\_template Sets whether or not a column for this locus is included in the Excel submission template.
  - Allowed: true/false (default: false)

#### Using existing locus definition as a template

When defining a new locus in the isolate database, it is possible to use an existing locus record as a template. To do this, click the 'Show tools' link in the top-right of the screen:

PubMLST / Users: Add	pecies home   Curator Query/update I   Query/update   Batch	's page (species)   Curator's page (d n insert	latabase)		
Logged in: Keith Jolley (keith). 🕒 Log o	ut   Change password				Toggle: 🚯
Add new locus					Show tools
Please fill in the fields below - req	uired fields are marke	d with an exclamation mark (!).			
Record				- Action	
id:1 data type:1 allele id format! length varies:1 coding sequence:1 flag table:1 isolate display:1 main display:1 query field:1 analysis:1	integer	Seqdef database supports allele flags		Reset Submit	
formatted name:			0		
common name: formatted common name:			0		
allele id regex:			Õ		
length:		<b>1</b>			
orf:	<b>v</b> 0				
genome position:		÷ 0			
-	🔘 true 🔘 false 🛈				
reference sequence:			ه اد		

This displays a drop-down box containing existing loci. Select the locus that you wish to use as a template, and click 'Copy'.

PTIDIALST / Users: Add	pecies home   Curator's page (species)   Curator's page (d Query/update d   Query/update   Batch insert	(database)
Logged in: Keith Jolley (keith). 🔂 Log o	ut   Change password	Toggle: 🜖
Add new locus		Hide tools
Please fill in the fields below - red	quired fields are marked with an exclamation mark (!).	Copy configuration from NEIS0001 (lpxC) Copy All parameters will be copied except Id, common name, reference sequence, genome position and length. The copied locus id will be substituted for "PUT_LOCUS_NAME_HERE"
id: data type: allele id formati length varies: coding sequence: flag table: isolate display. main display. query field. analysis:	DNA  integer  0 integer  0 itue  1 false 1 itue 1 i	in fields that include it.
formatted name: common name: formatted common name: allele id regex: length: off. genome position: match longest: reference sequence:	↓ 0 ↓ 0	0 0 0

The configuration will be copied over to the web form, with the exception of name fields. Some fields will require you to change the value 'PUT\_LOCUS\_NAME\_HERE' with the value you enter in the id field. These are usually the dbase\_id2\_value, description\_url and url fields:

allele id regex:		0	
length:	÷ 0	· ·	
orf:			
genome position:			
match longest:			
reference sequence:			
		0 6	
pcr filter:	🔿 true 🔘 false 🕦		
	◯ true ◯ false 0		
	pubmlst_bigsdb_neisseria_seqdef	Name of the database holding allele sequences	
dbase host:		1 IP address of database host	
dbase port:	🗧 🚺 Network port accepting database	connections	
dbase user:		0	
dbase password:		0	
dbase table:	sequences	Database table that holds sequence information for this locus	
dbase id field:		Primary field in sequence database that defines allele, e.g. 'allele_ic	d'
dbase id2 field:	locus	<ol> <li>Secondary field that defines allele, e.g. 'locus'</li> </ol>	
dbase id2 value:	PUT_LOCUS_NAME_HERE	<ol> <li>Secondary field value, e.g. locus name</li> </ol>	
dbase seq field:	sequence	Field in sequence database containing allele sequence	
description url:	/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_ne locus=PUT_LOCUS_NAME_HERE	isseria_seqdef&page=locusInfo&	
url:	/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst ne		
	<pre>locus=PUT_LOCUS_NAME_HERE&amp;allele_id=[?]</pre>		
		<b>0</b> h.	
submission template: aliases:	C true I false I Include column in isolate submission	template for this locus	
Action —			

Complete the form and click 'Submit'.

## **Batch adding**

Click the batch add (++) loci link on the curator's interface contents page.

		y/update   Ba	tormoon	
ged in: Keith Jolley (keith). 🗭 Log ou				
atabase curator's	inte	erface ·	<ul> <li>Neisseria</li> </ul>	PubMLST
Add, update or delet	te red	ords		
Record type	Add	Batch Add	Update or dele	te Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in th users table.
isolates	+	++	query   browse   batch update	
isolate field extended attribute values	+	++	?	Add values for additional isolate field attributes.
projects	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
isolate aliases	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags		scan	?	Tag regions of sequences within the sequence bin with locus information.
Database configurati Table		Batch Add L	pdate or delete	Comments
loci	+	++	?	
	data	bank scan		
locus aliases	+	++	? Add	alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++	Set	up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++	Def	ne nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
solate field extended attributes	+	++	? Def	ine additional attributes to associate with values of a particular isolate record field.

Click the link to download an Excel template:

Profit MLST Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert				
Logged in: Keith Jolley (keith). Log out   Change password			Toggle:	: <u>i</u>
Batch insert loci				
This page allows you to upload locus data as tab-delimited text or copied from a spreadsheet. <ul> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>Enter aliases (alternative names) for your locus as a semi-colon (;) separated list.</li> </ul>				
<ul> <li>Download tab-delimited header for your spreadsheet - use Paste special → text to paste the data.</li> <li>Download submission template (xisx format)</li> </ul>				
Paste in tab-delimited text (include a field header line).	Action			
	Reset	Submit		
Back				

Fill in the spreadsheet fields using the *table above as a guide*, then paste the completed table into the web form and press 'Submit query'.

# 5.14 Defining locus extended attributes

You may want to add additional metadata for the allele definitions of some loci. Since these are likely to be specific to each locus, they cannot be defined generically within the standard locus definition. We can, instead, define extended attributes. Examples of these include higher order grouping of antigen sequences, antibody reactivities, identification of important mutations, or cross-referencing of alternative nomenclatures.

To add extended attributes for a locus, click add (+) locus extended attributes in the sequence definition database curator's interface contents page.

user groups	+	++			Users can be members of these groups - use for setting access permissions.		
curator permissions				?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user table.		
locus descriptions	+	++		?			
scheme curator control lis	t +	++		query   batch	Define which curators can add or update profiles for particular schemes.		
locus curator control list	+	++		query   batch	Define which curators can add or update sequences for particular loci.		
sequences (all loci)	+	++   FAS	TA	?			
accession number links	+	++		?	Associate sequences with Genbank/EMBL accession number.		
PubMed links (to sequences)	+	++		?			
MLST profiles	+	++	query	browse   list   batch update			
PubMed links (to profiles)	+	++		?			
Table	Add	Batch Add	Update or delete		Comments		
loci	+	++	?				
		tabank scan					
locus aliases	+	++	?	Add alternative names	for loci. These can also be set when you batch add loci.		
ocus extended attributes	+	++	?	Define additional fields	to associate with sequences of a particular locus.		
client databases	+	++	?		hat use locus allele or scheme profile definitions defined in this database - this enables backlinks and searches n you query sequences or profiles in this database.		
client database loci	+	++	?	Define loci that are use	d in dient databases.		
lient database schemes	+	++	?	Define schemes that are	e used in client databases. You will need to add the appropriate loci to the client database loci table.		
client database fields linked to loci	+	++	?	Define fields in client d	Define fields in client database whose value can be displayed when isolate has matching allele.		
schemes	+	++	?	Describes schemes cons	sisting of collections of loci, e.g. MLST.		
scheme members	+	++	?	Defines which loci belo	ng to a scheme.		
scheme fields	+	++	?	Defines which fields be	long to a scheme.		
scheme groups	+	++	?	Describes groups in to v	hich schemes can belong - groups can also belong to other groups.		
scheme group scheme members	+	++	?	Defines which schemes belong to a group.			
scheme group group members	+	++	?	Defines which scheme (	groups belong to a parent group.		
0.000	- Se	ta user pa	assword to e	nable them to log on (	or change an existing password.		

Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

PubMLST)	Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley	keith). 🖗 Log out   Change password	Toggle: 🜖
Add new loo	cus extended attribute	
Please fill in the field	is below - required fields are marked with an exclamation mark (!).	
Record		Action
locus:	<b>•</b>	Reset Submit
field:		
value format:		
	© true ● false 0	
datestamp:	Keith Jolley (keith)	
value regex:	0	
description:	•	
option list:		
	h.	0
length:	·····	U
field order:		
nord order.	×	

- locus Select locus from dropdown box.
  - Allowed: existing locus name.
- field Name of extended attributes.
  - Allowed: any value.
- value\_format Data type of attribute.
  - Allowed: integer/text/boolean.
- required Specifies whether the attribute value but be defined when adding a new sequence.
  - Allowed: true/false.
- value\_regex Regular expression to enforce allele id naming (optional).
  - ^: the beginning of the string
  - \$:the end of the string
  - d: digit
  - D: non-digit
  - s: white space character
  - S: non white space character
  - w: alpha-numeric plus '\_'
  - .: any character
  - \*: 0 or more of previous character
  - +: 1 or more of previous character

- description Description that will appear within the web form when adding new sequences (optional).
  - Allowed: any value.
- option\_list Pipe (l) separated list of allowed values (optional).
- length Maximum length of value (optional).
  - Allowed: any integer.
- field\_order Integer that sets the position of the field within scheme values in any results (optional).
  - Allowed: any integer.

Once extended attributes have been defined, they will appear in the web form when adding new sequences for that locus. The values are searchable when using a *locus-specific sequence query*, and they will appear within query results and allele information pages.

# 5.15 Defining schemes

Schemes are collections of loci that may be associated with particular fields - one of these fields can be a primary key, i.e. a field that uniquely defines a particular combination of alleles at the associated loci.

A specific example of a scheme is MLST - see workflow for setting up a MLST scheme.

To set up a new scheme, you need to:

- 1. Add a new scheme description.
- 2. Define loci as 'scheme members'.
- 3. Add 'scheme fields' associated with the scheme.

#### See also:

```
Schemes (concept)
```

## 5.15.1 Sequence definition databases

As with all configuration, tables can be populated using the batch interface (++) or one at a time (+). Details for the latter are described below:

Click the add (+) scheme link on the curator's interface contents page.

MLST profiles			4	browse   list   batch update
PubMed links (to profile	s) +	++		?
Database con	figur	ation		
Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
		abank can		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
locus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlinks and searches of these databases when you query sequences or profiles in this database.
client database loci	+	++	?	Define loci that are used in client databases.
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
cheme group scheme members	+	++	?	Defines which schemes belong to a group.
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.
	ck - Cł	necks dat	tabase conne	enable them to log on or change an existing password. ctivity for loci and schemes and that required helper applications are properly installed.

Fill in the scheme description in the web form. The next available scheme id number will be filled in already.

The display\_order field accepts an integer that can be used to order the display of schemes in the interface - this can be left blank if you wish.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Add new scheme	
Please fill in the fields below - required fields are marked with an exclamation mark (!).  Record  id: 1  description:! MLST  curator:! Keith Jolley (keith)  datestamp:! 2014-07-04  date entered! 2014-07-04  date ontered! 2014-07-04  isplay order:  allow missing loci: © true © false [	Submit

To add loci to the scheme, click the add (+) scheme members link on the curator's interface contents page.

PubMed links (to profiles	s) +	++		2					
-ubmed links (to promes	9 T			7					
0									
Database cont	figur	ation							
Table	Add	Batch Add	Update or delete	Comments					
loci	+	++	?						
		abank can							
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.					
locus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.					
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlinks and searches of these databases when you query sequences or profiles in this database.					
client database loci	+	++	?	Define loci that are used in client databases.					
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.					
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.					
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.					
scheme members	+	++	?	Defines which loci belong to a scheme.					
scheme fields	+	++	?	Defines which fields belong to a scheme.					
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.					
cheme group scheme members	+	++	?	Defines which schemes belong to a group.					
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.					
	ck - Cl	necks da	tabase conne	enable them to log on or change an existing password. •ctivity for loci and schemes and that required helper applications are properly installed.					

Select the scheme name and a locus that you wish to add to the scheme from the appropriate drop-down boxes. *Loci need to have already been defined*. The field\_order field allows you to set the display order of the locus within a profile - if these are left blank that alphabetical ordering is used.

Database: Species home   Curator's page (specie Users: Add   Query/update Coc: Add MLST profiles: Add   Query/update   Batch insert	28)   Curator's page (database)	
Logged in: Keith Jolley (keith). Log out   Change password	Toggi	le: i
Add new scheme member		
Please be aware that any modifications to the structure of this scheme profiles will have to be reloaded.	will result in the removal of all data from it. This is done to ensure data integrity. This does not affect allele designations, bu	t any
Please fill in the fields below - required fields are marked with an excla	imation mark (!).	
Record	Action	
scheme id: MLST -	Reset Submit	
locus: abcZ		
curator:! Keith Jolley (keith)		
datestamp:! 2014-07-04 field order: 1		

To add scheme fields, click the add (+) scheme fields link on the curator's interface contents page.

MLST profiles			400171	browse   list   batch update				
PubMed links (to profile:	5) +	++		?				
Database con	figur	ation						
Table	Add	Batch Add	Update or delete	Comments				
loci	+	++	?					
		abank Ican						
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.				
locus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.				
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlinks and searches of these databases when you query sequences or profiles in this database.				
client database loci	+	++	?	Define loci that are used in client databases.				
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.				
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.				
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.				
scheme members	+	++	?	Defines which loci belong to a scheme.				
scheme fields	+	++	?	Defines which fields belong to a scheme.				
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.				
cheme group scheme members	+	++	?	Defines which schemes belong to a group.				
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.				
	ck - Cl	hecks da	tabase conne	enable them to log on or change an existing password. ctivity for loci and schemes and that required helper applications are properly installed.				

Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

PubMLST	Database: Species home   Curator's page (species)   Curator Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	r's page (database)	
Logged in: Keith Jolle	y (keith). Log out   Change password	Toggle:	: <u>i</u>
Add new s	cheme field		
Please be aware to profiles will have to		in the removal of all data from it. This is done to ensure data integrity. This does not affect allele designations, but a	any
Please fill in the fie	elds below - required fields are marked with an exclamation ma	ırk (!).	
Record		Action	
scheme id:!	MLST 🗸	Reset Submit	
field:!	ST		
	integer 👻		
	🖲 true 🔘 false 👔		
	🔿 true 🔘 false 👔		
	Keith Jolley (keith)		
datestamp:!	2014-07-04		
description:			
field order:	1 True O false i		
value regex:	i		

- scheme\_id Dropdown box of scheme names.
  - Allowed: selection from list.
- field Field name.
  - Allowed: any value.
- type Format for values.

- Allowed: text/integer/date.
- primary\_key Set to true if field is the primary key. There can only be one primary key for a scheme.
  - Allowed: true/false.
- dropdown Set to true if a dropdown box is displayed in the query interface, by default, for values of this field to be quickly selected. This option can be overridden by user preferences.
  - Allowed: true/false.
- description This field isn't currently used.
- field\_order Integer that sets the position of the field within scheme values in any results.
  - Allowed: any integer.
- value\_regex Regular expression to enforce field values.
  - ^: the beginning of the string
  - \$:the end of the string
  - d: digit
  - D: non-digit
  - s: white space character
  - S: non white space character
  - w: alpha-numeric plus '\_'
  - .: any character
  - \*: 0 or more of previous character
  - +: 1 or more of previous character

## 5.15.2 Isolate databases

As with all configuration, tables can be populated using the batch interface (++) or one at a time (+). Details for the latter are described below:

Click the add (+) scheme link on the curator's interface contents page.

sequences			2	The sequence on holds sequence contrigs from any source.				
accession number links	+	++		Associate sequences with Genbank/EMBL accession number.				
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.				
experiment sequences			?	Add links associating sequences to experiments.				
sequence tags	s	can	?	Tag regions of sequences within the sequence bin with locus information.				
Database configurat	tion							
Table	Add	Batch Add	Update or delete	Comments				
loci	+	++	?					
	datat	bank scan						
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.				
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.				
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.				
solate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.				
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.				
sequence attributes	+	++	?	Define attributes that can be set for contigs in the sequence bin.				
sets	+	++	?	Sets describe a collection of loci and schemes that can be treated like a stand-alone database.				
set loci	+	++	?	Add loci to sets.				
set schemes	+	++	?	Add schemes to sets.				
set metadata	+	++	?	Add metadata collection to sets.				
setview	+	++	?	Set database views linked to sets.				
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.				
scheme members	+	++	?	Defines which loci belong to a scheme.				
scheme fields	+	++	?	Defines which fields belong to a scheme.				
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.				
scheme group scheme members	+	++	?	Defines which schemes belong to a group.				
cheme group group members	+	++	?	Defines which scheme groups belong to a parent group.				
		•		to log on or change an existing password. and schemes and that required helper applications are properly installed.				

Fill in the scheme description in the web form. Required fields have an exclamation mark (!) next to them:

PubMLST / Users:	se: Species home   C Add   Query/update s: Add   Query/update	Curator's page (species)   Cur   Batch insert	ator's page (dat	abase)			
Logged in: Keith Jolley (keith). 🕩	Log out   Change passw	ord					Toggle: 🚯
Add new schem	е						
Please fill in the fields below	/ - required fields are	marked with an exclamation	mark (!).				
					-Action		
id:! 4	41				Reset	Submit	
description:			0		Reset	Submit	
isolate display:	🖲 true 🔘 false 🕦						
main display:! (	🖲 true 🔘 false 🚺						
query field:	🖲 true 🔘 false 🕦						
query status:	🛇 true 🔍 false 🚺						
analysis:! (	🖲 true 🔘 false 🚺						
curator:! K	(eith Jolley (keith)						
datestamp: 2	015-07-01						
date entered: 2	015-07-01						
dbase name:				0			
dbase host:			0				
dbase port:		÷ 0					
dbase user:			0				
dbase password:			0				
dbase table:			0				
display order:		<b>₽</b> 0					
allow missing loci:	🔿 true 🔘 false 🕦						

• id - Index number of scheme - the next available number will be entered automatically.

- Allowed: any positive integer.
- description Short description this is used in tables so make sure it's not too long.
  - Allowed: any text.
- isolate\_display Sets whether or not fields for this scheme are displayed in a detailed isolate record this can be overridden by user preference.
  - Allowed: allele only/sequence/hide.
- main\_display Sets whether or not fields for this scheme are displayed in a main results table by default this can be overridden by user preference.
  - Allowed: true/false.
- query\_field Sets whether or not fields for this scheme can be used in queries by default this can be overridden by user preference.

- Allowed: true/false.

- query\_status Sets whether a dropdown list box should be displayed in the query interface to filter results based on profile completion for this scheme this can be overridden by user preference.
  - Allowed: true/false.
- analysis Sets whether or not alleles for this locus can be used in analysis functions by default this can be overridden by user preference.
  - Allowed: true/false.
- dbase\_name Name of seqdef database (system name) containing scheme profiles (optional).
  - Allowed: any text.
- dbase\_host Resolved name of IP address of database host leave blank if running on the same machine as the isolate database (optional).
  - Allowed: network address, e.g. 129.67.26.52 or zoo-oban.zoo.ox.ac.uk
- dbase\_port Network port on which the sequence definition database server is listening leave blank unless using a non-standard port, 5432 (optional).
  - Allowed: integer.
- dbase\_user Name of user with permission to access the sequence definition database depending on the database configuration you may be able to leave this blank (optional).
  - Allowed: any text (no spaces).
- dbase\_password Password of database user again depending on the database configuration you may be able to leave this blank (optional).
  - Allowed: any text (no spaces).
- dbase\_table Table in the sequence definition database that contains profiles for this scheme. If the definition database uses BIGSdb this will be 'scheme\_X' where X is the scheme id number in the sequef database.
  - Allowed: any text (no spaces).
- display\_order Integer reflecting the display position for this scheme within the interface (optional).
  - Allowed: any integer.
- allow\_missing\_loci Allow profile definitions to contain '0' (locus missing) or 'N' (any allele).

# 5.16 Organizing schemes into hierarchical groups

Schemes can be organized in to groups, and these groups can in turn be members of other groups. This faciliates hierarchical ordering of loci, but with the flexibility to allow loci and schemes to belong to multiple groups.

This hierarchical structuring can be seen in various places within BIGSdb, for example the *allele download* page.

nip <u>M</u>	LST	Downl	oad: Al	ences   Batch sequences   Cor Ileles   MLST profiles nts   Home   PorA   FetA   Optio		ch profiles   List   Brow	se   Query	
								Help 🗗
ownlo	oad a	allele	e se	quences				
Select loci l	by scher	ne   Alpi	nabetic	al list   All loci by scheme				
				ails of loci belonging to schem 1e nodes to expand/collapse.	es or groups of schemes - clic	king a group folder wil	I display the lo	ci for all schemes within the
group and a	any sub <u>c</u>	proups.	CIICK II	te nodes to expand/collapse.				
🛓 📗 All I							<u> </u>	
	Capsule							
- 🔺 🛄 (	Genetic I	nformat	ion Pro	ocessing				
	🔲 Repli	cation a	nd Rep	pair				
	📗 DI	VA repli	cation					
	🗄 🛄 Ni	ucleotid	e excis	ion repair				
þ /	Trans	cription						
⊳	Trans	lation						
	Metaboli						=	
- 🗼 🛅 1	Typing							
	MLST							
	Finet		tigens					
	Antig	-	-					
		-		rtial genes)				
				nole genes)				
	Other scl		cus wi	iole genes)				
	oci not i						-	
	0.1101	in statet	nes					
MLST								
Locus D	ownload	I Type	Alleles	Length	Full name/product	Curator(s)	Last updated	
adk 🕕	*	DNA			test_full_name / test_product	O. Harrison, K. Jolley	2015-01-15	
abcZ 🚯	*	DNA	630	Variable: (432 min; 433 max)		O. Harrison, K. Jolley	2015-04-14	
aroE 🜖	±	DNA	682	Fixed: 490 bp		O. Harrison, K. Jolley	2013-01-22	
fumC 🚺	*	DNA	639	Fixed: 465 bp		O. Harrison	2013-01-23	
gdh 🕕	*	DNA	653	Fixed: 501 bp		O. Harrison	2013-01-22	
pdhC 🕕	*	DNA	652	Fixed: 480 bp		O. Harrison	2013-01-21	
pgm 📵	±	DNA	675	Fixed: 450 bp		O. Harrison	2013-01-22	
				L Encol format				
Download	table: tab	o-delimi	ted text	t   Excel format				

Scheme groups can be added in both the sequence definition and isolate databases. To add a new group, click the add (+) scheme group link on the curator's contents page.

++	?	Set up projects to which isolates can belong.
-+	?	Add isolates to projects.
++	?	Add alternative names for isolates.
++	?	
++	?	Allele designations can be set within the isolate table functions.
++	?	The sequence bin holds sequence contigs from any source.
++	?	Associate sequences with Genbank/EMBL accession number.
++	?	Set up experiments to which sequences in the bin can belong.
	?	Add links associating sequences to experiments.
scan	?	Tag regions of sequences within the sequence bin with locus information.
n d Batch Add	Update or delete	Comments
	?	
	?	Add alternative names for loci. These can also be set when you batch add loci.
		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
++	?	Define additional attributes to associate with values of a particular isolate record field.
	?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.
++	?	Define attributes that can be set for contigs in the sequence bin.
++	?	Sets describe a collection of loci and schemes that can be treated like a stand-alone database.
++	?	Add loci to sets.
++	?	Add schemes to sets.
++	?	Add metadata collection to sets.
++	?	Set database views linked to sets.
++	?	Describes schemes consisting of collections of loci, e.g. MLST.
++	?	Defines which loci belong to a scheme.
	?	Defines which fields belong to a scheme.
++		
++ ++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
	? ?	Describes groups in to which schemes can belong - groups can also belong to other groups. Defines which schemes belong to a group.
	+++            Scan            Scan            Batch Add            ++            Iabank scan            ++            ++            ++            ++            ++            ++            ++            ++            ++            ++            ++            ++            ++            ++            ++            ++            ++	++         ?           ++         ?           scan         ?           scan         ?           d         Batch Add         Update or delete           ++         ?           tabank scan

Enter a short name for the group - this will appear within drop-down list boxes and the hierarchical tree, so it needs to be fairly short.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert							
Logged in: Keith Jolley (keith). 🕪 Log out   Change password Toggle: 🚯							
Add new scheme group							
Please fill in the fields below - required fields are marked with an exclamation mark (!).  Record  id: 1  name: Typing  curator:! Keith Jolley (keith)  datestamp:! 2015-07-01  description:							
display order: seq query: O true O false Action Reset Submit							

If you are creating a scheme group in the sequence definition database, there is an additional field called 'seq\_query'. Set this to true to add the scheme group to the dropdown lists in the *sequence query* page. This enables all loci belonging to schemes within the group to be queried together.

Schemes can be added to groups by clicking the add (+) scheme group scheme members link.

Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
		tabank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
locus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlir and searches of these databases when you query sequences or profiles in this database.
client database loci	+	++	?	Define loci that are used in client databases.
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loc table.
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
scheme group scheme members	+	++	?	Defines which schemes belong to a group.
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.

Select the scheme and the group to add it to, then click 'Submit'.

Database: Species home   Curator's page (specie Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	is)   Curator's page (database)
Logged in: Keith Jolley (keith). Log out   Change password	Toggle:
Add new scheme group scheme memb	)er
Please fill in the fields below - required fields are marked with an exclar Record group id! Typing scheme id! MLST curator: Kenth Jolley (Kenth) datestamp:! 2014-07-10	mation mark (!). Action Reset Submit

Scheme groups can be added to other scheme groups in the same way by clicking the add (+) scheme group group members link.

# 5.17 Setting up client databases

Sequence definition databases can have any number of isolate databases that connect as clients. Registering these databases allows the software to perform isolate data searches relevant to results returned by the sequence definition database, for example:

- Determine the number of isolates that a given allele is found in and link to these.
- Determine the number of isolates that a given scheme field, e.g. a sequence type, is found in and link to these.
- Retrieve specific attributes of isolates that have a given allele, e.g. species that have a particular 16S allele, or penicillin resistance given a particular penA allele.

Multiple client databases can be queried simultaneously.

To register a client isolate database for a sequence definition database, click the add (+) client database link on the curator's interface contents page.

Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
		tabank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
locus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlini and searches of these databases when you query sequences or profiles in this database.
client database loci	+	++	?	Define loci that are used in client databases.
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
cheme group scheme members	+	++	?	Defines which schemes belong to a group.
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.

Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

Loc	tabase: Species home   Curator's page (species) ers: Add   Quer/lupdate i: Add ST profiles: Add   Quer/lupdate   Batch insert	) Curatori s (rage (database)		
Logged in: Keith Jolley (keith	<ul> <li>h). Log out   Change password</li> </ul>			Toggle: i
Add new clien	it database			
Please fill in the fields be	elow - required fields are marked with an exclama	ation mark (!).		
Record			Action	
id:	1		Reset Submit	
name:!	PubMLST isolates		Reset	
		f isolates that represent the total known		
	diversity of <u>Neisseria</u> species. H database there is at least one co	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be		
	diversity of <u>Neisseria</u> species. I database there is at least one co isolate may be submitted to this	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be		
dbase name:! dbase config name:!	diversity of Neisseria species. E database there is at least one co isolate may be submitted to this noted that it does not represent pubmlst_bigsdb_neisseria_isolates pubmlst_neisseria_isolates	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be a population sample.		
dbase name:! dbase config name:! curator:!	diversity of <u>Neisseria</u> species. E database there is at least one oc isolate may be submitted to this noted that it does not represent pubmlst_bigsdb_neisseria_isolates pubmlst_neisseria_isolates Keith Jolley (keith)	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be a population sample. Name of the database holding isolate data		
dbase name:1 dbase config name:1 curator1 datestamp:1	diversity of <u>Neisseria</u> species. I database there is at least one co isolate may be submitted to this noted that it does not represent pubmist_heisseria_isolates Keith Jolley (keith) 2014-07-04	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be a population sample. Name of the database holding isolate data Name of the database configuration		
dbase name!! dbase config name!! curator.! datestamp:! dbase host	diversity of Neisseria species. I database there is at least one co isolate may be submitted to this noted that it does not represent pubmlst_bigsdb_neisseria_isolates pubmlst_neisseria_isolates Keith Jolley (keith) 2014-07-04	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be a population sample. Name of the database holding isolate data Name of the database configuration		
dbase name:! dbase config name:! curator.! datestamp:! dbase host dbase port	diversity of Neisseria species. I database there is at least one co isolate may be submitted to this noted that it does not represent pubmlst_bigsdb_neisseria_isolates pubmlst_neisseria_isolates Keith Jolley (keith) 2014-07-04	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be a population sample. Name of the database holding isolate data Name of the database configuration		
dbase name:! dbase config name:! curator.! datestamp:! dbase host dbase port. dbase user:	diversity of Neisseria species. E database there is at least one co isolate may be submitted to this noted that it does not represent pubmlst_bigsdb_neisseria_isolates pubmlst_neisseria_isolates Keith Jolley (keith) 2014-07-04	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be a population sample. Name of the database holding isolate data Name of the database configuration		
dbase name! dbase config name! curator! datestamp! dbase host dbase user. dbase user. dbase user.	diversity of Neisseria species. I database there is at least one co isolate may be submitted to this noted that it does not represent pubmist_bigsdb_neisseria_isolates pubmist_neisseria_isolates Keith Jolley (keith) 2014-07-04	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be a population sample. Name of the database holding isolate data Name of the database configuration [] IP address of database host pling database connections [] ]		
dbase name! dbase config name! curator! datestamp: dbase host dbase port dbase user: dbase password dbase view:	diversity of Neisseria species. I database there is at least one co isolate may be submitted to this noted that it does not represent pubmist_bigsdb_neisseria_isolates pubmist_neisseria_isolates Keith Jolley (keith) 2014-07-04	For every allelic profile in the profiles orresponding isolate deposited here. Any database and consequently it should be a population sample. Name of the database holding isolate data Name of the database configuration		

- id Index number of client database. The next available number is entered automatically but can be overridden.
  - Allowed: any positive integer.
- name Short description of database. This is used within the interface result tables so it is better to make it as short as possible.

- Allowed: any text.
- description Longer description of database.
  - Allowed: any text.
- dbase\_name Name of database (system name).
  - Allowed: any text.
- dbase\_config\_name Name of database configuration this is the text string that appears after the db= part of script URLs.
  - Allowed: any text (no spaces)
- dbase\_host Resolved name of IP address of database host (optional).
  - Allowed: Network address, e.g. 129.67.26.52 or zoo-oban.zoo.ox.ac.uk
  - Leave blank if running on the same machine as the sequef database.
- dbase\_port Network port on which the client database server is listening (optional).
  - Allowed: integer.
  - Leave blank unless using a non-standard port (5432).
- dbase\_user Name of user with permission to access the client database.
  - Allowed: any text (no spaces).
  - Depending on the database configuration you may be able to leave this blank.
- · dbase\_password Password of database user
  - Allowed: any text (no spaces).
  - Depending on the database configuration you may be able to leave this blank.
- url URL of client database bigsdb.pl script
  - Allowed: valid script path.
  - This can be relative (e.g. /cgi-bin/bigsdb/bigsdb.pl) if running on the same machine as the seqdef database or absolute (including http://) if on a different machine.

## 5.17.1 Look up isolates with given allele

To link a locus, click the add (+) client database loci link on the curator's interface contents page.

Table	Add	Batch Add	Update or delete	Comments		
loci	+	++	?			
	da	tabank				
	scan					
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.		
locus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.		
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlin		
				and searches of these databases when you query sequences or profiles in this database.		
client database loci	+	++	?	Define loci that are used in client databases.		
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.		
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.		
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.		
scheme members	+	++	?	Defines which loci belong to a scheme.		
scheme fields	+	++	?	Defines which fields belong to a scheme.		
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.		
cheme group scheme members	+	++	?	Defines which schemes belong to a group.		
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.		

Link the locus to the appropriate client database using the dropdown list boxes. If the locus is named differently in the client database, fill this name in the locus\_alias.

PubMLST La	atabase: Species home   Curator's page (spec sers: Add   Query/update ci: Add .ST profiles: Add   Query/update   Batch insert			
Logged in: Keith Jolley (ke	th). Log out   Change password			Toggle: i
Add new locu	is to client database defi	inition		
Please fill in the fields	below - required fields are marked with an exc	lamation mark (!).	Action	
	1) PubMLST isolates 👻		Reset Submit	
locus:!				
curator:! datestamp:!	Keith Jolley (keith) 2014-07-04			
locus alias:		name that this locus is referred by in client database (if different)		

Now when information on a given allele is shown following a query, the software will list the number of isolates with that allele and link to a search on the database to retrieve these.

(PubMLST) D	tuery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Jownload: Alleles   MLST profiles inks: Contents   Home   PorA   FetA   Options   Isolate Database
Allele inform	ation - abcZ: 5
Provenance/meta	a data
locus:	abcZ
allele:	5
sequences:	TTIGATACCE TIGCCGAAGE TITGEGCGAA ATTCGCGATI TATTGCGCCE TIATCATCAT GICAGCCAIG AGTIGGAAAA IGGTICGAGI GAGGCTIIGI IGAAAGAGT TAACGAAIG CAACIIGAA ICGAAGCCAA GGACGGCTGG AAGCIGGATG CGGCAGICAA GCAGACITIG GGIGGACITG GITGGCCAGA AAACGAAAAA ATCGGCAACC ICTCCGGCGG ACAGAAAAA CGIGIGCCC IAGGGCAGGC IIGGGIGCAG AAGCCIGAIG IATIGCIGCI GGACGAACCCA ACCAACCAII IGGACAIIGA CGCGAITAII IGGCIGGAAA AICIGCITAA AGCGIIIGAA GGCAGCCIGG IIGIGAIIAC CCAGGACCGC CGIIIIIIGG ACAAIAICCC CAGCGCAIC GICGAACICG AIC
length:	433
status:	Sanger trace checked
date entered:	2001-02-07
datestamp:	2009-11-11
sender:	Keith Jolley, University of Oxford, UK
curator:	Man-Suen Chan, University of Oxford (E-mail: man-suen.chan@paediatrics.ox.ac.uk)
Profiles containin	ng this allele
MLST:	183 profiles
Isolate databases	5
PubMLST isolates:	Contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

## 5.17.2 Look up isolates with a given scheme primary key

Setting this up is identical to setting up for alleles (see above) except you click on the add (+) client database schemes link and choose the scheme and client databases in the dropdown list boxes.

Now when information on a given scheme profile (e.g. MLST sequence type) is shown following a query, the software will list the number of isolates with that profile and link to a search on the database to retrieve these.



## 5.17.3 Look up specific isolate database fields linked to a given allele

To link an allele to an isolate field, click the add (+) 'client database fields linked to loci' link on the curator's interface contents page.

Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
		tabank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
locus extended attributes	+	++	?	Define additional fields to associate with sequences of a particular locus.
client databases	+	++	?	Add isolate databases that use locus allele or scheme profile definitions defined in this database - this enables backlink and searches of these databases when you query sequences or profiles in this database.
client database loci	+	++	?	Define loci that are used in client databases.
client database schemes	+	++	?	Define schemes that are used in client databases. You will need to add the appropriate loci to the client database loci table.
client database fields linked to loci	+	++	?	Define fields in client database whose value can be displayed when isolate has matching allele.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
cheme group scheme members	+	++	?	Defines which schemes belong to a group.
scheme group group members	+	++	?	Defines which scheme groups belong to a parent group.

Select the client database and locus from the dropdown lists and enter the isolate database field that you'd like to link. The 'allele\_query' field should be set to true.

Database: Species home   Curator's page (species)   Curator's page (databas PutplALST) Users: Add   Querylupdate Loci: Add	se)
MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Add new locus to client database isolate field definition	on
Please fill in the fields below - required fields are marked with an exclamation mark (!).	
Record	— — Action —
client dbase id: 1) PubMLST isolates 👻	Reset Submit
locus:! penA	
isolate field: penicillin_range	
curator:! Keith Jolley (keith)	
datestamp:! 2014-07-04	
allele query:	2.

Now, in the allele record or following a sequence query that identifies an allele, all values for the chosen field from isolates with the corresponding allele are shown.

ele morm	
	ation - penA: 9
ovenance/meta	a data
locus:	penA
allele:	9
sequences:	GACGECGTIT IGCIGCCGGI CAGCTITGAA AAACAGGCGG IIGCGCCGCA AGGCAAACGI AIAITIAAAG CAICGACCGC ACGICAGGIG CGIGAGIIGA IGGIIICIGA AACCAGAACCI GGCGGIACG GIACGGCGGG IGCGGIAGAI GGIIICGACC IGGCGCAAA AACCGGIACG GCGCGIAAGI IGGIIAACGG ICGIIACGIC GAIIACAAC ACGIIGCCAC IIICAICGGI IIIGCCCCGG CIAAAAAICC GCGIGIGAII GGIGCGGIAA CAIIGACGA GCCGACIGCA AACGGIIACI ACGGCGGCGI AGIGACAGGI CCGGICTICA AACAAGIIAI GGGCGGIAGC CIGAACAICI IGGGCGIIIC ICCGACCAAA CCICIGACCA AI
length:	402
status:	Sanger trace checked
date entered:	2006-09-04
datestamp:	2006-09-04
sender:	Muhamed-Kheir Taha, National Reference Center for Meningococci, Institut Pasteur, France
curator:	Keith Jolley, University of Oxford, UK (E-mail: keith.jolley@zoo.ox.ac.uk)
mutation F504L:	yes
mutation A510V:	yes
mutation I515V:	yes
mutation H541N:	
mutation I566V:	yes
blication (1)	
Jolley KA, Kad	uez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, Cafferkey MT, Carion F, Christensen JJ, Diggle M, Edwards G, Enríquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S lubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen L, Levenet I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel L L (2007). Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis. <i>Antimicrob Agents Chemother</i> 51:2784-92
olate databases	5
ubMLST isolates:	Contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample. 141 isolates
nked data	
incu uata	

# 5.18 Rule-based sequence queries

The RuleQuery plugin has been designed to extract information from a pasted-in genome sequence, look up scheme fields and client database fields, and then format the output in a specified manner.

Rules are written in Perl, allowing the full power of this scripting language to be utilised. Helper functions that perform specific actions are available to the script (see example).

Please note that direct access to the database is prevented as are system calls.

## 5.18.1 Example rule code

An example can be found on the Neisseria sequence database that takes a genome sequence and determines a fine type and antibiotic resistance.

The code for this rule is as follows:

```
#Clinical identification rule
#Update job viewer status
update_status({stage=>'Scanning MLST loci'});
#Scan genome against all scheme 1 (MLST) loci
scan_scheme(1);
```

```
#Update job viewer status
update_status({percent_complete=>30, stage=>'Scanning PorA and FetA VRs'});
#Scan genome against the PorA VR and FetA VR loci
scan_locus($_) foreach qw(PorA_VR1 PorA_VR2 FetA_VR);
Add text to main output
append_html("<h1>Strain type</h1>");
#Set variables for the scanned results. These can be found in the
#$results->{'locus'} hashref
my %alleles;
$alleles{$_} = $results->{'locus'}->{$_} // 'ND' foreach qw(PorA_VR1 PorA_VR2);
$alleles{'FetA_VR'} = $results->{'locus'}->{'FetA_VR'} // 'F-ND';
#Scheme field values are automatically determined if a complete
#profile is available. These are stored in the $results->{'scheme'} hashref
my $st = $results->{'scheme'}->{1}->{'ST'} // 'ND';
append_html("P1.$alleles{'PorA_VR1'}, $alleles{'PorA_VR2'}; $alleles{'FetA_VR'}; ST-$st ");
#Reformat clonal complex using a regular expression, e.g.
#'ST-11 clonal complex/ET-37 complex' gets rewritten to 'cc11'
my $cc = $results->{'scheme'}->{1}->{'clonal_complex'} // '-';
$cc =~ s/ST-(\S+) complex.*/cc$1/;
append_html("($cc)");
if ($st eq 'ND') {
 append_html("ST not defined. If individual MLST loci have been found "
  . "they will be displayed below:");
 #The get_scheme_html function automatically formats output for a scheme.
 #Select whether to display in a table rather than a list, list all loci, and/or list fields.
 append_html(get_scheme_html(1, {table=>1, loci=>1, fields=>0}));
}
#Antibiotic resistance
update_status({percent_complete=>80, stage=>'Scanning penA and rpoB'});
scan_locus($_) foreach qw(penA rpoB);
if (defined $results->{'locus'}->{'penA'} || defined $results->{'locus'}->{'rpoB'} ){
 append html("<h1>Antibiotic resistance</h1>");
 if (defined $results->{'locus'}->{'penA'}){
   append_html("<i>penA</i> allele: $results->{'locus'}->{'penA'}");
    #If a client isolate database has been defined and values have been defined in
    #the client_dbase_loci_fields table, the values for a field in the isolate database can be
    #retrieved based on isolates that have a particular allele designated.
    #The min_percentage attribute states that only values that are represented by at least that
    #proportion of all isolates that had a value set are returned (null values are ignored).
   my $range = get_client_field(1, 'penA', 'penicillin_range', {min_percentage => 75});
   append_html(" (penicillin MIC: $range->[0]->{'penicillin_range'})") if @$range;
   append_html("");
 if (defined $results->{'locus'}->{'rpoB'}){
   append_html("<i>rpoB</i> allele: $results->{'locus'}->{'rpoB'}");
   my $range = get_client_field(1, 'rpoB', 'rifampicin_range', {min_percentage => 75});
   append_html(" (rifampicin MIC: $range->[0]->{'rifampicin_range'})") if @$range;
    append_html("");
  }
```

append\_html("");

#### **Rule files**

The rule file is placed in a rules directory within the database configuration directory, e.g. /etc/bigsdb/dbase/pubmlst\_neisseri\_seqdef/rules. Rule files are suffixed with '.rule' and their name should be descriptive since it is used within the interface, i.e. the above rule file is named Clinical\_identification.rule (underscores are converted to spaces in the web interface).

#### Linking to the rule query

Links to the rule query are not automatically placed within the web interface. The above rule query can be called using the following URL:

http://pubmlst.org/perl/bigsdb/bigsdb.pl?db=pubmlst\_neisseria\_seqdef&page=plugin&name=RuleQuery&ruleset=Clinical\_identification

To place a link to this within the database contents page an HTML file called job\_query.html can be placed in a contents directory within the database configuration directory, e.g. in /etc/bigsdb/dbases/pubmlst\_neisseria\_seqdef/contents/job\_query.html. This file should contain a list entry (i.e. surrounded with and

#### Adding descriptive text

Descriptive text for the rule, which will appear on the rule query page, can be placed in a file called description.html in a directory with the same name as the rule within the rule directory, e.g. in /etc/bigsdb/dbases/pubmlst\_neisseria\_seqdef/rules/Clinical\_identification/description.html.

## 5.19 Workflow for setting up a MLST scheme

The workflow for setting up a MLST scheme is as follows (the example seqdef database is called seqdef\_db):

#### Seqdef database

- 1. Create appropriate loci
- 2. Create new scheme 'MLST'
- 3. Add scheme\_field 'ST' with primary\_key=TRUE (add clonal\_complex if you want; set this with primary\_key=FALSE)
- 4. Add each locus as a scheme\_member
- 5. You'll then be able to add profiles

#### Isolate database

- 1. Create the same loci with the following additional parameters (example locus 'atpD')
- dbase\_name: seqdef\_db
- dbase\_table: sequences
- dbase\_id\_field: allele\_id
- dbase\_id2\_field: locus

- dbase\_id\_value: atpD
- dbase\_seq\_field: sequence
- url: something like /cgi-bin/bigsdb/bigsdb.pl?db=seqdef\_db&page=alleleInfo&locus=atpD&allele\_id=[?]
- 2. Create scheme 'MLST' with:
- dbase\_name: seqdef\_db
- dbase\_table: scheme\_1 (or whatever the id of your seqdef scheme is)
- 3. Add scheme\_field ST as before
- 4. Add loci as scheme\_members

# 5.20 Defining new loci based on annotated reference genome

An annotated reference genome can be used as the basis of defining loci. The 'Databank scan' function will create an upload table suitable for pasting directly in to the batch locus add form of the *sequence definition* or *isolate* databases.

allele designations sequences		++	?	Allele designations can be set within the isolate table functions. The sequence bin holds sequence contigs from any source.					
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.					
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.					
experiment sequences			?	Add links associating sequences to experiments.					
sequence tags	s	scan	?	Tag regions of sequences within the sequence bin with locus information.					
S Database configura	tion								
Table	Add	Batch Add	Update or delete	Comments					
loci	+	++	?						
	data	bank scan							
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.					
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.					
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specifi repetitive loci.					
olate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.					
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.					
sequence attributes	+	++	?	Define attributes that can be set for contigs in the sequence bin.					
sets	+	++	?	Sets describe a collection of loci and schemes that can be treated like a stand-alone database.					
set loci	+	++	?	Add loci to sets.					
set schemes	+	++	?	Add schemes to sets.					
set metadata	+	++	?	Add metadata collection to sets.					
setview	+	++	?	Set database views linked to sets.					
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.					
scheme members	+	++	?	Defines which loci belong to a scheme.					
scheme fields	+	++	?	Defines which fields belong to a scheme.					
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.					
scheme group scheme members	+	++	?	Defines which schemes belong to a group.					
heme group group members	+	++	?	Defines which scheme groups belong to a parent group.					
				n to log on or change an existing password. ci and schemes and that required helper applications are properly installed.					

Click 'Database scan' on the curator's contents pag.

Enter an EMBL or Genbank accession number for a complete annotated genome and press 'Submit'.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert
Logged in: Keith Jolley (keith). Log out   Change password
Scan EMBL/Genbank record for loci
This function allows you to scan an EMBL or Genbank (whole genome) file in order to create a batch upload file for setting up new loci.  Please enter accession number Primary identifier Accession AM421808  O gene name Submit

A table of loci will be generated provided a valid accession number is provided.

РарМІ	.ST/υ	atabase: Species home   Curator's page (species)   Curator's page (database) sers: Add   Query/update plates: Add   Query/update   Batch insert	
Logged in: Keit	h Jolley (ke	ith). Log out   Change password	
Scan El	MBL/C	Senbank record for loci	
This function	n allows yo	u to scan an EMBL or Genbank (whole genome) file in order to create a batch upload file for setting up new loci.	
Please e	enter acce	ssion number — — Primary identifier - — Action —	
Accession:	AM4218	8 O locus tag Submit	
		© gene name	
Download tab	ole: tab-de	imited text   Excel format (suitable for batch upload of loci).	
Download all	eles: tab-o	elimited text   Excel format (suitable for defining the first allele in the seqdef database).	
Annotatio	n inform	ation	
	cession: version:	AM421808	
	type:		
		2194961	
des	-	Veisseria meningitidis serogroup C FAM18 complete genome.	
	regions:		
Coding se	equence	S	
Locus	Aliases	Product	Length
NMC0001	IpxC;	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase	924
NMC0002	envA pilS1	pilin (fragment)	291
NMC0002	pilS2	prin (regiment) truncated pilin	366
NMC0004	fbp	peptidyl-prolyl cis-trans isomerase	330
NMC0005		putative membrane protein	219
NMC0006		putative glycerate dehydrogenase	954
NMC0007	metG	methionyl-tRNA synthetase	2058
NMC0008	glmS	glucosaminefructose-6-phosphate aminotransferase [isomerizing]	1839
NMC0009		putative lipoprotein	519
NMC0010	gna33	outer membrane lipoprotein Gna33 []	1326
NMC0011		putative integral membrane protein	840
NMC0012 NMC0013			4467
		putative lipoprotein	1167
NMC0014	nhnA	putative lipoprotein possible membrane protein	1266
NMC0014 NMC0015	phnA glmU	putative lipoprotein	

Tab-delimited text and Excel format files will be created to be used as the basis for upload files for the sequence definition and isolate databases. Batch sequence files, in text and Excel formats, are also created for defining the first allele once the locus has been set up in the sequence definition database.

and all all the second	Is	plates: Add   Query/update   Batch insert	
ogged in: Keitl	h Jolley (ke	th). Log out   Change password	
can El	MBL/G	Genbank record for loci	
This function	allows yo	u to scan an EMBL or Genbank (whole genome) file in order to create a batch upload file for setting up new loci.	
— Please e Accession:		8 Olocus tag Gene name	
		imited text   Excel format (suitable for batch upload of loci). elimited text   Excel format (suitable for defining the first allele in the seqdef database).	
	version:		
des	type: length:	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome.	
des coding Coding se	type: length: ccription: regions: equence	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5	Lengt
des coding coding se Locus NMC0001	type: length: ccription: regions: equence Aliases lpxC; envA	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 S Product UDP-3-O-[3-hydroxymyristoy]] N-acetylglucosmine deacetylase	Leng 924
des coding oding se Locus NMC0001 NMC0002	type: length: ccription: regions: equence Aliases lpxC; envA pilS1	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 S Product UDP-3-O-[3-hydroxymyristoy]] N-acetylglucosmine deacetylase pillin (fragment)	924 291
des coding se Locus NMC0001 NMC0002 NMC0003	type: length: ccription: regions: equence Aliases lpxC; envA pilS1 pilS2	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5 <b>Product</b> UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase pilin (fragment) truncated pilin	924 291 366
des coding coding se Locus NMC0001 NMC0002 NMC0003 NMC0004	type: length: ccription: regions: equence Aliases lpxC; envA pilS1	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5 <b>Product</b> UDP-3-O-[3-hydroxymyristoyi] N-acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-prolyl cis-trans isomerase	924 291 366 330
des coding oding se Locus VMC0001 VMC0002 VMC0003 VMC0004 VMC0005	type: length: ccription: regions: equence Aliases lpxC; envA pilS1 pilS2	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 S Product UDP-3-O-[3-hydroxymyristoy]] N-acetylglucosmine deacetylase UDP-3-O-[3-hydroxymyristoy]] N-acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-prolyl cis-trans isomerase putative membrane protein	924 291 366 330 219
des coding oding se Locus NMC0001 NMC0002 NMC0003 NMC0003 NMC0005 NMC0006	type: length: cription: regions: equence Aliases IpxC; envA piIS1 piIS2 fbp	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 S Product UDP-3-O-[3-hydroxymyristoy]] N-acetylglucosmine deacetylase UDP-3-O-[3-hydroxymyristoy]] N-acetylglucosmine deacetylase pilin (fragment) truncated pilin putative glycerate dehydrogenase	924 291 366 330 219 954
des coding se coding se Locus NMC0001 NMC0002 NMC0003 NMC0004 NMC0005 NMC0005 NMC0005 NMC0007	type: length: cription: regions: equence Aliases lpxC; envA pilS1 pilS2 fbp metG	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5 <b>Product</b> UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-prolyl cis-trans isomerase putative membrane protein putative glycerate dehydrogenase methionyl-IRNA synthetase	924 291 366 330 219 954 2058
des coding se coding se Locus NMC0001 NMC0002 NMC0003 NMC0004 NMC0005 NMC0006 NMC0006 NMC0008	type: length: cription: regions: equence Aliases IpxC; envA piIS1 piIS2 fbp	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5 <b>Product</b> UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase IDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase I	924 291 366 330 219 954 2058 1839
des coding se coding se Locus NMC0001 NMC0003 NMC0003 NMC0005 NMC0005 NMC0006 NMC0009	type: length: cription: regions: equence Aliases lpxC; envA pilS1 pilS2 fbp metG glmS	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 s Product UDP-3-O-[3-hydroxymyristoy[] N-acetylglucosmine deacetylase UDP-3-O-[3-hydroxymyristoy[] N-acetylglucosmine deacetylase pilin (fragment) truncated pilin putative glucosamine-fructose-6-phosphate aminotransferase [isomerizing] putative lipoprotein	924 291 366 330 219 954 2056 1839 519
des coding se coding se coding se NMC0001 NMC0002 NMC0003 NMC0005 NMC0005 NMC0008 NMC0008 NMC0009 NMC0009 NMC0010	type: length: cription: regions: equence Aliases lpxC; envA pilS1 pilS2 fbp metG	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5 5 6 7 7 7 7 7 7 7 7 7 7 7 7 7	924 291 366 330 219 954 2058 1839 519 1326
des coding se coding se Locus NMC0001 NMC0003 NMC0004 NMC0006 NMC0006 NMC0008 NMC0009 NMC0000 NMC0011	type: length: cription: regions: equence Aliases lpxC; envA pilS1 pilS2 fbp metG glmS	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5 5 6 7 7 7 7 7 7 7 7 7 7 7 7 7	924 291 366 330 219 954 2058 1839 519 1326 840
des coding se Locus NMC0001 NMC0002 NMC0005 NMC0006 NMC0007 NMC0007 NMC0008 NMC0010 NMC0012	type: length: cription: regions: equence Aliases lpxC; envA pilS1 pilS2 fbp metG glmS	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5 5 7 7 7 7 7 7 7 7 7 7 7 7 7	_
des coding se Locus NMC0001 NMC0002 NMC0003 NMC0006 NMC0006 NMC0006 NMC0000 NMC0000 NMC0000 NMC0011	type: length: cription: regions: equence Aliases lpxC; envA pilS1 pilS2 fbp metG glmS	dna 2194961 Neisseria meningitidis serogroup C FAM18 complete genome. 1975 5 5 6 7 7 7 7 7 7 7 7 7 7 7 7 7	924 291 360 330 219 954 205 183 519 132 840 116

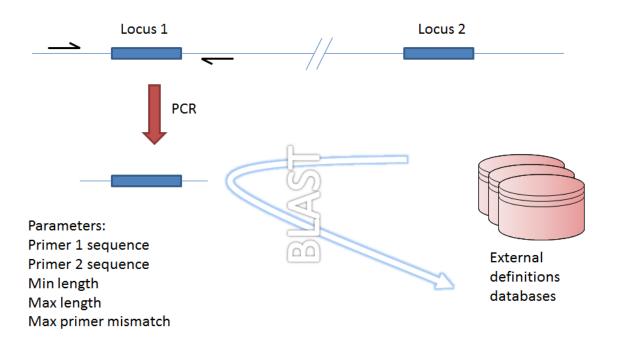
# 5.21 Genome filtering

Within a genome there may be multiple loci that share allele pools. If an allele sequence is tagged from a genome using only BLAST then there is no way to determine which locus has been identified. It is, however, possible to further define loci by their context, i.e. surrounding sequence.

## 5.21.1 Filtering by in silico PCR

Provided a locus can be predicted to be specifically amplifed by a PCR reaction, the genome can be filtered to only look at regions prediced to fall within amplification products of one or more PCR reactions. Since this is *in silico* we don't need to worry about problems such as sequence secondary structure and primers can be any length.

To define a PCR reaction that can be linked to a locus definition, click the add (+) PCR reaction link on the curator's main page.



Locus 1 and locus 2 share allele pool

Fig. 5.1: Genome filtering by in silico PCR.

allolo designations		++	2	Allele designations can be set within the isolate table functions						
allele designations		++	2	Allele designations can be set within the isolate table functions.						
sequences				The sequence bin holds sequence contigs from any source.						
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.						
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.						
experiment sequences			?	Add links associating sequences to experiments.						
sequence tags	S	scan	?	Tag regions of sequences within the sequence bin with locus information.						
Database configura	ation									
Table	Add	Batch Add	Update or delete	Comments						
loci	+	++	?							
	datat	bank scan								
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.						
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetiti loci.						
nucleotide probes	+	++		Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci.						
solate field extended attributes	; +	++	?	Define additional attributes to associate with values of a particular isolate record field.						
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.						
sequence attributes	+	++	?	Define attributes that can be set for contigs in the sequence bin.						
sets	+	++	?	Sets describe a collection of loci and schemes that can be treated like a stand-alone database.						
set loci	+	++	?	Add loci to sets.						
set schemes	+	++	?	Add schemes to sets.						
set metadata	+	++	?	Add metadata collection to sets.						
setview	+	++	?	Set database views linked to sets.						
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.						
scheme members	+	++	?	Defines which loci belong to a scheme.						
scheme fields	+	++	?	Defines which fields belong to a scheme.						
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.						
scheme group scheme members	+	++	?	Defines which schemes belong to a group.						
cheme group group members	s +	++	?	Defines which scheme groups belong to a parent group.						
		-		n to log on or change an existing password. ci and schemes and that required helper applications are properly installed.						

In the resulting web form you can enter values for your two primer sequences (which can be any length), the minimum and maximum lengths of reaction products you wish to consider and a value for the allowed number of mismatches per primer.

PTIDMLST / Users:	se: Species home   Curator's page (species)   Curator's page (database) Add   Query/update : Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). 🗭	Log out   Change password	Toggle: 🜖
Add new PCR re	action	
Please fill in the fields below	- required fields are marked with an exclamation mark (!).	
Record		_
id:!	1	
description:		
primer1:		
primer2:		
	Keith Jolley (keith)	
datestamp: min length:	2015-07-01	
max length:		
max primer mismatch:		
Action		
Reset Submit		

- id PCR reaction identifier number.
  - Allowed: integer.
- description Description of PCR reaction product.
  - Allowed: any text.
- primer1 Primer 1 sequences
  - Allowed: nucleotide sequence (IUPAC ambiguous characters allowed).
- primer2 Primer 2 sequence.
  - Allowed: nucleotide sequence (IUPAC ambiguous characters allowed).
- min\_length Minimum length of predicted PCR product.
  - Allowed: integer.
- max\_length Maximum length of predicted PCR product.
- max\_primer\_mismatch Number of mismatches allowed in primer sequence.
  - Allowed: integer.
  - Do not set this too high or the simulation will run slowly.

Associating this with a particular locus is a two step process. First, create a locus link by clicking the add (+) PCR locus link on the curator's main page. This link will only appear once a PCR reaction has been defined.

experiments	+	++	?	Set up experiments to which sequences in the bin can belong.							
experiment sequences			?	Add links associating sequences to experiments.							
sequence tags	S	scan	?	Tag regions of sequences within the sequence bin with locus information.							
Database configura	ation										
Table	Add	Batch Add	Update or delete	Comments							
loci	+	++	?								
	databank scan										
locus aliases	+ ++ ?		?	Add alternative names for loci. These can also be set when you batch add loci.							
PCR reactions	+	++	?	Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.							
PCR locus links	+	++		Link a locus to an in silico PCR reaction.							
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.							
olate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.							
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.							
sequence attributes	+	++	?	Define attributes that can be set for contigs in the sequence bin.							
sets	+	++	?	Sets describe a collection of loci and schemes that can be treated like a stand-alone database.							
set loci	+	++	?	Add loci to sets.							
set schemes	+	++	?	Add schemes to sets.							
set metadata	+	++	?	Add metadata collection to sets.							
setview	+	++	?	Set database views linked to sets.							
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.							
scheme members	+	++	?	Defines which loci belong to a scheme.							
scheme fields	+	++	?	Defines which fields belong to a scheme.							
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.							
scheme group scheme members	+	++	?	Defines which schemes belong to a group.							
cheme group group members	; +	++	?	Defines which scheme groups belong to a parent group.							
				n to log on or change an existing password. ci and schemes and that required helper applications are properly installed.							

Select the locus and PCR reaction name from the dropdown lists to create the link. You also need to edit the locus table and set the pcr\_filter field to 'true'.

Now when you next perform tag scanning there will be an option to use PCR filtering.

## 5.21.2 Filtering by in silico hybridization

An alternative is to define a locus by proximity to a single probe sequence. This is especially useful if you have multiple contigs and the locus in question may be at the end of a contig so that it doesn't have upstream or downstream sequence available for PCR filtering.

The process is very similar to setting up PCR filtering, but this time click the nucleotide probe link on the curator's content page.

Enter the nucleotide sequence and a name for the probe. Next you need to link this to the locus in question. Click the add (+) probe locus links link on the curator's main page. This link will only appear once a probe has been defined.

Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

- probe\_id Dropdown list of probe names.
  - Allowed: selection from list.
- locus Dropdown list of loci.
  - Allowed: selection from list.
- max\_distance Minimum distance of probe from end of locus.

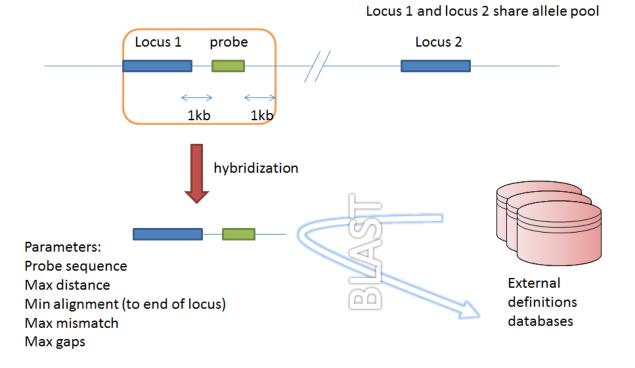


Fig. 5.2: Filtering by in silico hybridization

- Allowed: any positive integer.
- min\_alignment Minimum length of alignment allowed.
  - Allowed: any positive integer.
- max\_mismatch Maximum number of mismatches allowed in alignment.
  - Allowed: any positive integer.
- max\_gaps Maximum number of gaps allowed in alignment.
  - Allowed: any positive integer.

Finally edit the locus table and set the probe\_filter field for the specified locus to 'true'.

Now when you next perform tag scanning there will be an option to use probe hybridization filtering.

# 5.22 Setting locus genome positions

The genome position for a locus can be set directly by editing the locus record. To batch update multiple loci based on a tagged genome, however, a much easier way is possible. For this method to work, the reference genome must be represented by a single contig.

From the curator's main page, you need to do a query to find the isolate that you will base your numbering on. Click 'isolate query' to take you to a standard query form.

accession number links	•		?	Associate sequences with Genbank/EMBL accession number.						
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.						
experiment sequences			?	Add links associating sequences to experiments.						
sequence tags	S	can	?	Tag regions of sequences within the sequence bin with locus information.						
Database configura	tion	Datah	lindata ar							
Table	Add	Batch Add	Update or delete	Comments						
loci	+ ++ databank scan		?							
locus aliases	+ ++		?	Add alternative names for loci. These can also be set when you batch add loci.						
PCR reactions	+	++	?	Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.						
PCR locus links	+	++		Link a locus to an <i>in silico</i> PCR reaction.						
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.						
solate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.						
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.						
sequence attributes	+	++	?	Define attributes that can be set for contigs in the sequence bin.						
sets	+	++	?	Sets describe a collection of loci and schemes that can be treated like a stand-alone database.						
set loci	+	++	?	Add loci to sets.						
set schemes	+	++	?	Add schemes to sets.						
set metadata	+	++	?	Add metadata collection to sets.						
setview	+	++	?	Set database views linked to sets.						
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.						
scheme members	+	++	?	Defines which loci belong to a scheme.						
scheme fields	+	++	?	Defines which fields belong to a scheme.						
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.						
scheme group scheme members	+	++	?	Defines which schemes belong to a group.						
cheme group group members		++	2	Defines which scheme groups belong to a parent group.						

Set user passwords - Set a user password to enable them to log on or change an existing password.
 Configuration check - Checks database connectivity for loci and schemes and that required helper applications are properly installed.

experiments	+	++	?	Set up experiments to which sequences in the bin can belong.						
experiment sequences			?	Add links associating sequences to experiments.						
sequence tags	S	can	?	Tag regions of sequences within the sequence bin with locus information.						
Database configura	ition									
Table	Add	Batch Add	Update or delete	Comments						
loci	+	++	?							
	databank scan									
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.						
PCR reactions	+	++	?	Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.						
PCR locus links	+	++		Link a locus to an <i>in silico</i> PCR reaction.						
nucleotide probes	+	++	?	Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.						
probe locus links	+	++		Link a locus to an in silico hybridization reaction.						
plate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.						
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.						
sequence attributes	+	++	?	Define attributes that can be set for contigs in the sequence bin.						
sets	+	++	?	Sets describe a collection of loci and schemes that can be treated like a stand-alone database.						
set loci	+	++	?	Add loci to sets.						
set schemes	+	++	?	Add schemes to sets.						
set metadata	+	++	?	Add metadata collection to sets.						
setview	+	++	?	Set database views linked to sets.						
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.						
scheme members	+	++	?	Defines which loci belong to a scheme.						
scheme fields	+	++	?	Defines which fields belong to a scheme.						
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.						
scheme group scheme members	+	++	?	Defines which schemes belong to a group.						
heme group group members	+	++	?	Defines which scheme groups belong to a parent group.						

Configuration check - Checks database connectivity for loci and schemes and that required helper applications are properly installed.

Database: Species home | Curator's page (species) | Curator's page (database) Users: Add | Query/update PubMLST Isolates: Add | Query/update | Batch insert

Logged in: Keith Jolley (keith). DLog out | Change password

experiments

## Database curator's interface - Neisseria PubMLST

Record type	Add	Batch Add	Update or delete	Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a stat 'curator' in the users table.
isolates	+	++	query browse   list   batch update	
solate field extended attribute values	+	++	?	Add values for additional isolate field attributes.
projects	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
isolate aliases	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	2	Associate sequences with Genbank/EMBL accession number.

0

Set up experiments to which sequences in the bin can belong

#### Perform your search and click the hyperlinked id number of the record.

++

+

Publ	MLST	💟 Users: A	dd   Query	/update	urator's pa Batch ins		es) C	urator's p	page (database)										
Logged in:	: Keith Jol	lley (keith). Lo	g out   Chan	ige password	1							Toggle: i	Field	help: id					Go
Isolat	te qu	ery/up	date																
Isol	ate prove	enance/phen	otype field	s					— Display/sort o	ptions ——									Modify
isola	ate	•	=	<b>→</b> M	58		+	• 1	Order by: Display:		ds p	er page <u>i</u>	▼ asc	ending	•				form options
Dele	returned.	Submit . Click the hyp — — Tag sc	perlink for anning –	Projec	s														
Delet	te ALL	Sca	in	Select p	oject			•	Link				- Eine 4					40	
Delete	Update	Sequence bin	New version	id isola	ite aliases			fields <u>i</u> disease		serogroup	ST	MLST clonal complex	PorA VR1	ping an PorA VR2	FetA VR	test ST	tes PorA VR1	PorA VR2	Loci 'test9
delete	update	upload	create	240 MC	58 Z7176	UK	1983		Neisseria meningitidis	В	74	ST-32 complex/ET-5 complex	7 update	16-2 update	F1-5 update		7 update	16-2 update	add

In the isolate record, click the sequence bin 'Display' button to bring up details of the isolate contigs.

species: Neisseria meningiti	dis update history: 137 updates show details
Publications (4)	
	r, Quail MA, Cherevach I, White B, Parkhill J, Maiden MC (2010). Independent evolution of the core and accessory gene sets in the genus genome of Neisseria lactamica isolate 020-06. BMC Genomics 11:652 79 isolates
	ton C, Bentley SD, Parkhill J, Maiden MC (2012). A genomic approach to bacterial taxonomy: an examination and proposed reclassification of Microbiology 158:1570-80 55 isolates
<ul> <li>Stabler RA, Marsden GL, Witney AA, Neisseria species. <i>Microbiology</i> 15<sup>-</sup></li> </ul>	Li Y, Bentley SD, Tang CM, Hinds J (2005). Identification of pathogen-specific genes through microarray analysis of pathogenic and commensal 1:2907-22 3 isolates
Haft DH, Salzberg SL, White O, Fleis	J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Hickey EK, chmann RD, Dougherty BA, Mason T, Ciecko A, Parksey DS, Blair E, Cittone H, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Vamathevan J, M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC (2000). Complete genome sequence of Neisseria meningitidis 87:1809-15 Isolate
Sequence bin	
contigs: 1 length: 2272360 bp loci tagged: 1283	detailed breakdown: Display
Schemes and loci	
All loci     Capsule     Genetic Information Processing     All Action     Genetic Information Processing     All Action     Genetic Information     Genetic Information     All Action     Genetic Information     Genet	Navigate and select schemes within tree to display allele designations

### Click the 'Renumber' button:

sequence method designation Length Comments test accession attribute date int Locus Start End Direction format 1	ged in: Keith	n Jolley (keith). Log ou	ut   Change password	i								Toggle:
<ul> <li>Sequence sequencing Original Length Comments float test accession attribute date int Locus Start End Direction EMBL Artemis Renum format attribute date int NEIS2140 502 897 ← EMBL Artemis Renum format attribute date int NEIS2140 502 897 ← EMBL Artemis Renum format attribute date int NEIS2141 918 2312 ← NEIS2141 918 2312 ← NEIS2141 918 2312 ← NEIS2143 3158 3511 ← NEIS2143 3158 3511 ← NEIS2144 3635 4117 → NEIS2144 3635 4117 → NEIS2146 4958 5975 → NEIS2146 4958 6214 → NEIS2146 4958 6214 → NEIS2147 5936 6214 → NEIS2151 10350 10811 ← NEIS2150 9346 10317 ← NEIS2151 10350 10811 ← NEIS2150 9346 10317 ← NEIS2150 9346 10317 ← NEI</li></ul>	auen	ce bin for l	MC58									
<ul> <li>Number of contigs: 1         <ul> <li>Length: 2272360</li> <li>Download sequences (FASTA format)</li> <li>Download sequences with annotations (EMBL format)</li> </ul> </li> <li>Sequence Sequencing Original Length Comments float test accession attribute date int Locus Start End Direction EMBL format end designation Length Comments test accession attribute date int Locus Start End Direction EMBL format end end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end end designation (EMBL Artemis Renum end end end end end end end end end end</li></ul>					 							
<ul> <li>Number of contigs: 1         <ul> <li>Length: 2272360</li> <li>Download sequences (FASTA format)</li> <li>Download sequences with annotations (EMBL format)</li> </ul> </li> <li>Sequence Sequencing Original Length Comments float test accession attribute date int Locus Start End Direction EMBL format end designation Length Comments test accession attribute date int Locus Start End Direction EMBL format end end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end end designation Length Comments test accession attribute date int Locus Start End Direction EMBL Artemis Renum end end designation (EMBL Artemis Renum end end end end end end end end end end</li></ul>	ntia sum	mary statistics										
<ul> <li>Length: 2272360</li> <li>Download sequences (FASTA format)</li> <li>Download sequences with annotations (EMBL format)</li> </ul> Sequence Sequencing Original designation Length Comments float test accession attribute date int Locus Start End Direction EMBL Artemis Renum genome 1 Sanger 2272360 whole genome is a sequence genome is a sequence genome is a sequence is a sequence is a sequence of the sequence is a sequen	-	1										
<ul> <li>Download sequences (FASTA format)</li> <li>Download sequences with annotations (EMBL format)</li> <li>Sequence Sequencing Original designation Length Comments test test test accession attribute date tint Locus Start End Direction Format Provide Artemis Renum Prevanting Preva</li></ul>												
• Download sequences with annotations (EMBL format)          Sequence       Sequencing method       Original designation       Length       Comments       float test accession       test attribute       test date       test int       Locus       Start       End       Direction       EMBL format       Artemis       Renum         1       Sanger       2272360       whole genome       accession       attribute       attribute       attribute       float       Start       End       Direction       EMBL format       Artemis       Renum         1       Sanger       2272360       whole genome       accession       attribute       attribute       float       Start       End       Direction       EMBL format       Artemis       Renum         1       Sanger       2272360       whole genome       accession       attribute       float       start       float       Start       End       EMBL format       Artemis       Renum         NEIS2142       2517       3161       ←       NEIS2143       3158       3511       ←       NEIS2143       3158       5875       →       NEIS2145       NEIS2146       4958       5875       →       NEIS2149       7573       8826       ←       NEIS2150       9346       10317	-		074 (1999)									
Sequence sequencing method designation       Length       Comments       float test accession       test attribute       test date       test int       Locus       Start       End       Direction       EMBL format       Artemis       Renum         1       Sanger       2272360       whole genome       whole genome       accession       attribute       test       test       test       test       test       nels2140       502       897       EMBL       Artemis       Renum         1       Sanger       2272360       whole genome       accession       attribute       test       test       nels2141       918       2312       EMBL       Artemis       Renum         NEIS2142       2517       3161       Comments       istart       Nels2142       2517       3161       Comments       Nels2142       Nels2142       2517       3161       Comments       Nels2142       Nels2145 <td< td=""><td></td><td></td><td></td><td>(BL format)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>				(BL format)								
Sequence       method       designation       Length       Comments       test       accession       attribute       date       int       Locus       Start       End       Direction       format       I         1       Sanger       2272360       whole genome       accession       attribute       date       int       Locus       Start       End       Direction       format       I       Artemis       Renum         1       Sanger       2272360       whole genome       accession       attribute       date       int       NEIS2140       502       897       C       EMBL       Artemis       Renum         NEIS2142       2517       3161       C       NEIS2143       3158       3511       C       NEIS2144       3635       4117       Artemis       NEIS2145       4311       4961       Artemis       NEIS2145       4311       4961       Artemis       NEIS2145       4358       5875       Artemis       NEIS2145       4358       5875       Artemis       NEIS2149       7573       8826       C       NEIS2149       7573       8826       C       NEIS2152       10360       10811       C       NEIS2152       10840       12177       K(kdA)       K(kdA	• Dowin	oau sequences wit	n annotations (En	ibe formaty								
Sequence       method       designation       Length       Comments       test       accession       attribute       date       int       Locus       Start       End       Direction       format       I         1       Sanger       2272360       whole genome       accession       attribute       date       int       Locus       Start       End       Direction       format       I       Artemis       Renum         1       Sanger       2272360       whole genome       accession       attribute       date       int       NEIS2140       502       897       C       EMBL       Artemis       Renum         NEIS2142       2517       3161       C       NEIS2143       3158       3511       C       NEIS2144       3635       4117       Artemis       NEIS2145       4311       4961       Artemis       NEIS2145       4311       4961       Artemis       NEIS2145       4358       5875       Artemis       NEIS2145       4358       5875       Artemis       NEIS2149       7573       8826       C       NEIS2149       7573       8826       C       NEIS2152       10360       10811       C       NEIS2152       10840       12177       K(kdA)       K(kdA												
genome       NEIS2141       918       2312       ←         NEIS2142       2517       3161       ←         NEIS2143       3158       3511       ←         NEIS2144       3635       4117       →         NEIS2145       4311       4961       →         NEIS2146       4958       5875       →         NEIS2146       6281       7492       ←         (pgk)       NEIS2151       10317       ←         NEIS2151       10350       10811       ←         NEIS2151       10840       12177       ←         (kdA)       12177       ←	equence			Comments			Locus	Start	End	Direction		 Renumber
$\begin{array}{c ccccc} NEIS2142 & 2517 & 3161 & \longleftarrow \\ NEIS2143 & 3158 & 3511 & \longleftarrow \\ NEIS2144 & 3635 & 4117 & \rightarrow \\ NEIS2145 & 4311 & 4961 & \rightarrow \\ NEIS2146 & 4958 & 5875 & \rightarrow \\ NEIS2146 & 4958 & 5875 & \rightarrow \\ NEIS2147 & 5936 & 6214 & \rightarrow \\ NEIS2148 & 6281 & 7492 & \longleftarrow \\ (ngt) & & & & \\ NEIS2149 & 7573 & 8826 & \longleftarrow \\ NEIS2150 & 9346 & 10317 & \longleftarrow \\ NEIS2151 & 10350 & 10811 & \longleftarrow \\ NEIS2152 & 10840 & 12177 & \longleftarrow \\ NEIS2152 & 10840 & 12$	1	Sanger	2272360	) whole			NEIS2140	502	897	←	EMBL	 Renumber
$\begin{array}{c ccccc} NEIS2143 & 3158 & 3511 & \longleftarrow \\ NEIS2144 & 3635 & 4417 & \longrightarrow \\ NEIS2145 & 4311 & 4961 & \longrightarrow \\ NEIS2146 & 4958 & 5875 & \longrightarrow \\ NEIS2146 & 6281 & 7492 & \longleftarrow \\ (ngk) & & & & & \\ NEIS2149 & 7573 & 8826 & \longleftarrow \\ NEIS2150 & 9346 & 10317 & \longleftarrow \\ NEIS2151 & 10350 & 10811 & \longleftarrow \\ NEIS2152 & 10840 & 12177 & \longleftarrow \\ NEIS2152 & 10$				genome						←		
$\begin{array}{c ccccc} NEIS2144 & 3635 & 4117 & \longrightarrow \\ NEIS2145 & 4311 & 4961 & \longrightarrow \\ NEIS2146 & 4958 & 5875 & \longrightarrow \\ NEIS2147 & 5936 & 6214 & \longrightarrow \\ NEIS2148 & 6281 & 7492 & \longleftarrow \\ (pgk) & & & & & & & \\ NEIS2149 & 7573 & 8826 & \longleftarrow \\ NEIS2149 & 7573 & 8826 & \longleftarrow \\ NEIS2151 & 10350 & 10317 & \longleftarrow \\ NEIS2151 & 10350 & 10811 & \longleftarrow \\ NEIS2152 & 10840 & 12177 & \longleftarrow \\ (kdtA) & & & & & & & \\ \end{array}$										←		
$\begin{array}{c cccc} NEIS2145 & 4311 & 4961 & \longrightarrow \\ NEIS2146 & 4958 & 5875 & \longrightarrow \\ NEIS2147 & 5936 & 6214 & \longrightarrow \\ NEIS2148 & 6281 & 7492 & \longleftarrow \\ (pgk) & & & & & \\ NEIS2149 & 7573 & 8826 & \longleftarrow \\ NEIS2150 & 9346 & 10317 & \longleftarrow \\ NEIS2151 & 10350 & 10811 & \longleftarrow \\ NEIS2151 & 10350 & 10811 & \longleftarrow \\ NEIS2152 & 10840 & 12177 & \longleftarrow \\ (kdtA) & & & & & & \\ \end{array}$												
$\begin{array}{c cccc} NEIS2146 & 4958 & 5875 & \longrightarrow \\ NEIS2147 & 5936 & 6214 & \longrightarrow \\ NEIS2148 & 6281 & 7492 & \longleftarrow \\ (pgk) & & & & & \\ NEIS2149 & 7573 & 8826 & \longleftarrow \\ NEIS2150 & 9346 & 10317 & \longleftarrow \\ NEIS2151 & 10350 & 10811 & \longleftarrow \\ NEIS2152 & 10840 & 12177 & \longleftarrow \\ NEIS2152 & 10840 & 12177 & \longleftarrow \\ (kdtA) & & & & & \\ \end{array}$												
$\begin{array}{c cccc} NEIS2147 & 5936 & 6214 & \longrightarrow \\ NEIS2148 & 6281 & 7492 & \longleftarrow \\ (pgk) & & & & & \\ NEIS2149 & 7573 & 8826 & \longleftarrow \\ NEIS2150 & 9346 & 10317 & \longleftarrow \\ NEIS2151 & 10350 & 10811 & \longleftarrow \\ NEIS2152 & 10840 & 12177 & \longleftarrow \\ (kdtA) & & & & & \\ \end{array}$												
NEIS2148 (pgk)     6281     7492     ()       NEIS2149     7573     8826     ()       NEIS2150     9346     10317     ()       NEIS2151     10350     10811     ()       NEIS2152     10840     12177     ()												
(pgk) NEIS2149 7573 8826 ← NEIS2150 9346 10317 ← NEIS2151 10350 10811 ← NEIS2152 10840 12177 ← (kdtA)												
NEIS2149     7573     8826       NEIS2150     9346     10317       NEIS2151     10350     10811       NEIS2152     10840     12177       (kdtA)     10840     12177								6281	/492	←		
NEIS2150       9346       10317       ←         NEIS2151       10350       10811       ←         NEIS2152       10840       12177       ←         (kdtA)       10400       12177       ←								7570	0006			
NEIS2151 10350 10811 ← NEIS2152 10840 12177 ← (kdtA)												
NEIS2152 10840 12177 ← (kdtA)												
(KdtA)												
NEIS2153 12174 13622 🔶							(kdtA)			<i>—</i>		
NEI90001 15221 16144 -							NEIS2153	12174	13622	←		

A final confirmation screen is displayed with the option to remove existing numbering that doesn't appear within the reference genome. Click 'Renumber'.

PubMLST	Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert							
Logged in: Keith Jolle	e <b>y (keith)</b> . Log out   Change pa	assword						
Renumber	locus denom	e nositions ba	sed on tagged sequences					
Renativer	loouo genonii							
You have colocted	to repumber the gapome	positions set in the locus	table based on the tagged sequences in sequence id#1.					
rou nave selected	i to renumber the genome	positions set in the locus	table based on the tagged sequences in sequence id# i.					
Option		Actic	n					
Remove pos	itions for loci not tagged in	this sequence Reni	umber					
The following des	ignations will be made:							
	sting genome position 🔶	New genome position 🔶						
NEIS2140	2181973	502						
NEIS2140	2182389	918						
NEIS2142	2183899	2517						
NEIS2143	2184552	3158						
NEIS2144	2185029	3635						
NEIS2145	2185705	4311						
NEIS2146	2186352	4958						
NEIS2147	2187330	5936						
NEIS2148	2187675	6281						
NEIS2149	2188934	7573						
NEIS2150	2190551	9346						
NEIS2151	2191558	10350						
NEIS2152	2192048	10840						
NEIS2153	2193382	12174						
NEIS0001	1261	15221						
NEIS0210	209923	17229						
pilS	3271	18127						
NEIS0004	4069	23904						
NEIS0005	4476	24311						
NEIS0007	5843	25679						

# 5.23 Defining composite fields

Composite fields are virtual fields that don't themselves exist within the database but are made up of values retrieved from other fields or schemes and formatted in a particular way. They are used for display and analysis purposes only and can not be searched against.

One example of a composite field is used in the Neisseria PubMLST database which has a strain designation composite field made up of serogroup, PorA VR1 and VR2, FetA VR, ST and clonal complex designations in the format:

[serogroup]: P1.[PorA\_VR1],[PorA\_VR2]: [FetA\_VR]: ST-[ST] ([clonal\_complex])

e.g. A: P1.5-2,10: F1-5: ST-4 (cc4)

Additionally, the clonal complex field in the above example is converted using a regular expression from 'ST-4 complex/subgroup IV' to 'cc4'.

Composite fields can be added to the database by clicking the add (+) composite fields link on the curator's main page.

Database configura	tion			
Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
	datal	bank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
olate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.
sequence attributes	+	++		Define attributes that can be set for contigs in the sequence bin.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
scheme group scheme members	+	++	?	Defines which schemes belong to a group.
heme group group members	+	++	?	Defines which scheme groups belong to a parent group.
				m to log on or change an existing password. ci and schemes and that required helper applications are properly installed.

Initially you just enter a name for the composite field and after which field it should be positioned. You can also set whether or not it should be displayed by default in main results tables following a query - this is overrideable by user preferences.

PubMLST U	atabase: Species hon sers: Add   Query/upd olates: Add   Query/up		tor's page (database)		
Logged in: Keith Jolley (ke	eith). Log out   Change pa	ssword			Toggle: i
Add new con	nposite field	1			
Record	below - required fields	s are marked with an exclamation m	ark (!).	Action	
position after:!		<ul> <li>field present in the isolate table</li> </ul>		Reset Submit	
	◯ true	Sets whether to display field in isolate of	query results table (can be overridden by user preference).		

Once the field has been created it needs to be defined. This can be done from query composite field link on the main curator's page.

Se Database configura	tion			
Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
	datat	bank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci.
olate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.
sequence attributes	+	++		Define attributes that can be set for contigs in the sequence bin.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
scheme group scheme members	+	++	?	Defines which schemes belong to a group.
cheme group group members	+	++	?	Defines which scheme groups belong to a parent group.
				m to log on or change an existing password. ci and schemes and that required helper applications are properly installed.

### Select the composite field from the list and click 'Update'.

PubMLST	Database: Spec Users: Add   Que Isolates: Add   Q	ery/update		e (species)   Curator's page (database) t				
Logged in: Keith Jolle	ogged in: Keith Jolley (keith). Log out   Change password							
Update or	delete con	nposite	e field					
1 composite field c	lefined.	<b>pecitie</b>	main					
Delete Update	field name	position after		definition	missing data			
	field name strain_designation	after isolate	display false	definition [serogroup]: P1.[PorA_VR1].[PorA_VR2]: [FetA_VR]: ST-[scheme 1:ST] ([scheme 1:clonal_complex])	missing data ND: P1.ND,ND: F-ND: ST-ND (-)			

From this page you can build up your composite field from snippets of text, isolate field, locus and scheme field values. Enter new values in the boxes at the bottom of the page.

PubMLST / Users: A	dd   Query/u	iome   Curator's page (species)   pdate /update   Batch insert	Guraron's page (Garabase)
ogged in: Keith Jolley (keith). Log	out   Change	password	
Ipdate composi	te field	- strain_designat	ion
position after: isolate		-	
main display: false 👻 curator: Keith Jolley			
datestamp: 2014-04-09		Update	
field e	empty value	regex	curator datestamp delete edit move
Serogroup [isolate field]	ND		Keith Jolley 2009-11-12 delete edit up down
: P1.			Keith Jolley 2009-11-12 delete edit up down
PorA_VR1 [locus]	ND		Keith Jolley 2009-11-12 delete edit up down
,			Keith Jolley 2009-11-12 delete edit up down
PorA_VR2 [locus]	ND		Keith Jolley 2009-11-12 delete edit up down
:			Keith Jolley 2009-11-12 delete edit up down
FetA_VR [locus]	F-ND		Keith Jolley 2009-11-12 delete edit up down
: ST-			Keith Jolley 2009-11-12 delete edit up down
ST [MLST field]	ND		Keith Jolley 2009-11-12 delete edit up down
(			Keith Jolley 2009-11-12 delete edit up down
clonal_complex [MLST field]	-	s/ST-(\S+) complex.*/cc\$1	/ Keith Jolley 2009-11-12 delete edit up down
)			Keith Jolley 2009-11-12 delete edit up down
Add new field:			
		Add new text field	
	-	Add new isolate field	
	-	Add new locus field	
	•	Add new scheme field	
	curator	: Keith Jolley	
	datestamp	: 2014-07-08	

Once a field has been added to the composite field, it can be edited by clicking the 'edit' button next to it to add a regular expression to modify its value by specific rules, e.g. in the clonal complex field above, the regular expression is set as:

s/ST-(\S+) complex.\*/cc\$1/

which extracts one or more non-space characters following the 'ST-' in a string that then contains the work 'complex', and appends this to 'cc' to produce the final string.

This will convert 'ST-4 complex/subgroup IV' to 'cc4'.

You can also define text to be used for when the field value is missing, e.g. 'ND'.

# 5.24 Extended provenance attributes (lookup tables)

Lookup tables can be associated with an isolate database field such that the database can be queried by extended attributes. An example of this is the relationship between continent and country - every country belongs to a continent but you wouldn't want to store the continent with each isolate record (not only could data be entered inconsistently but it's redundant). Instead, each record may have a country field and the continent is then determined from the lookup table, allowing, for example, a search of isolates limited to those from Europe.

To set up such an extended attribute, click the add (+) isolate field extended attributes link on the curator's main page.

Database configura	ition			
Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
	datal	bank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci.
olate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.
sequence attributes	+	++		Define attributes that can be set for contigs in the sequence bin.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
scheme group scheme members	+	++	?	Defines which schemes belong to a group.
heme group group members	+	++	?	Defines which scheme groups belong to a parent group.
				m to log on or change an existing password. ci and schemes and that required helper applications are properly installed.

Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

- isolate\_field Dropdown list of isolate fields.
  - Allowed: selection from list.
- attribute Name of extended attribute, e.g. continent.
  - Allowed: any text (no spaces).
- value\_format Format for values.
  - Allowed: integer/float/text/date.
- value\_regex Regular expression to enforce allele id naming.
  - ^: the beginning of the string
  - \$:the end of the string
  - d: digit
  - D: non-digit
  - s: white space character
  - S: non white space character
  - w: alpha-numeric plus '\_'
  - .: any character
  - \*: 0 or more of previous character
  - +: 1 or more of previous character
  - e.g. ^Fd-d+\$ states that a value must begin with a F followed by a single digit, then a dash, then one or more digits, e.g. F1-12
- description Long description this isn't currently used but may be in the future.
  - Allowed: any text.

- url URL used to hyperlink values in the isolate information page. Instances of [?] within the URL will be substituted with the value.
  - Allowed: any valid URL (either relative or absolute).
- length Maximum length of extended attribute value.
  - Allowed: any positive integer.
- field\_order Integer that sets the order of the field following it's parent isolate field.
  - Allowed: any integer.

The easiest way to populate the lookup table is to do a batch update copied from a spreadsheet. Click the batch add (++) isolate field extended attribute values link on the curator's main page (this link will only appear once an extended attribute has been defined).

Database: Spe Users: Add   Qu Isolates: Add   Qu	Jery/up	date	tor's page (species)   Curator's tch insert	page (database)
gged in: Keith Jolley (keith). ⊕Log out	Change	e password		
atabase curator's i	nter	face -	Neisseria PubN	II ST
	mu	nucc -	ncioocna i aom	
Add, update or delete	reco	rds		
Record type	Add	Batch Add	Update or delete	Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table
isolates	+	++	query   browse   list   batch up	
olate field extended attribute value	s +	++	?	Add values for additional isolate field attributes.
projects	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
isolate aliases	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags		scan	?	Tag regions of sequences within the sequence bin with locus information.
Database configuratio		itch Add U	pdate or delete	Comments
loci	+	++	?	
		nk scan		
locus aliases	+	++		names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		de probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
isolate field extended attributes	+	++		al attributes to associate with values of a particular isolate record field.
composite fields	+			t composite fields consisting of fields from isolate, loci or scheme fields.
sequence attributes	+	++		s that can be set for contigs in the sequence bin.
schemes			2 Describes schem	nes consisting of collections of loci, e.g. MLST.

Download the Excel template:

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert			
Logged in: Keith Jolley (keith). Log out   Change password			Toggle: i
Batch insert isolate value extended attributes			
This page allows you to upload isolate field extended attribute value data as tab-delimited text or copied from a spreadshee • Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish. • Download tab-delimited header for your spreadsheet - use Paste special → text to paste the data. • Download submission template (xlsx format)	t.		
Paste in tab-delimited text (include a field header line).	Action -		
	Reset	Submit	
h.			
Back			

Fill in the columns with your values, e.g.

isolate_field	attribute	field_value	value
country	continent	Afghanistan	Asia
country	continent	Albania	Europe
country	continent	Algeria	Africa
country	continent	Andorra	Europe
country	continent	Angola	Africa

Paste from the spreadsheet in to the upload form and click 'Submit'.

# 5.25 Sequence bin attributes

It is possible that you will want to store extended attributes for sequence bin contigs when you upload them. Examples may be read length, assembler version, etc. Since there are almost infinite possibilities for these fields, and they are likely to change over time, they are not hard-coded within the database. An administrator can, however, create their own attributes for a specific database and these will then be available in the web form when uploading new contig data. The attributes are also searchable.

To set up new attributes, click the add (+) 'sequence attributes' link on the isolate database curator's index page.

Database configura	tion			
Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
	datat	bank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
olate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.
sequence attributes	+	++		Define attributes that can be set for contigs in the sequence bin.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
scheme group scheme members	+	++	?	Defines which schemes belong to a group.
cheme group group members	+	++	?	Defines which scheme groups belong to a parent group.
				m to log on or change an existing password. ci and schemes and that required helper applications are properly installed.

Enter the name of the attribute as the 'key', select the type of data (text, integer, float, date) and an optional short description. Click 'Submit'.

Database: Species home   Curator's page (species)   Cur Users: Add   Query/update Isolates: Add   Query/update   Batch insert	ator's page (database)	
Logged in: Keith Jolley (keith). Log out   Change password		Toggle: i
Add new sequence attribute		
Please fill in the fields below - required fields are marked with an exclamation Record	mark (!). — Action —	
key:! read_length type:! integer ✔	Reset Submit	
curator:: Keith Joliey (Keith)	1	
datestamp: 2014-07-15		
description:		

This new attribute will then be available when uploading contig data.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert				
Logged in: Keith Jolley (keith). Log out   Change password				Toggle: i
Batch insert sequences				
This page allows you to upload sequence data for a specified isolate record in FASTA format.				
If an isolate id is chosen, then all sequences will be associated with that isolate. Alternatively, the isolate id, or any other iso FASTA file. This allows data for multiple isolates to be uploaded.	late table field that uniquely	y defines the isolate, can be n	amed in the identifier ro	ows of the
Please note that you can reach this page for a specific isolate by querying isolates and then clicking "Upload" within the isola	te table.			
Please fill in the following fields - required fields are marked with an exclamation mark (!).				
Paste in sequences in FASTA format	- Attributes			
		Read identifier from FASTA	-	
	identifier field:			
	sender: !		-	
	method: run id:	•		
	assembly id:		_	
	read length:			
	- Options			
		ences shorter than 250 👻	bps.	
	Link to experiment:			
		FASTA file — Action –		
	Select FASTA file:	Reset	Submit	
	Browse No file se			
Back				

# 5.26 Checking external database configuration settings

Click the 'Configuration check' link on the curator's index page.

Table	Add	Batch Add	Update or delete	Comments			
loci	+	++	?				
	databank scar						
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.			
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.			
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.			
plate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.			
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.			
sequence attributes	+	++		Define attributes that can be set for contigs in the sequence bin.			
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.			
scheme members	+	++	?	Defines which loci belong to a scheme.			
scheme fields	+	++	?	Defines which fields belong to a scheme.			
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.			
scheme group scheme members	+	++	?	Defines which schemes belong to a group.			
heme group group members	+	++	?	Defines which scheme groups belong to a parent group.			

The software will check that required helper applications are installed and executable and, in isolate databases, test every locus and scheme external database to check for connectivity and that data can be retrieved.

ged in: Keith Jolley (k	keith). Log out   Change password										
onfiguratio	on check - Neisseria P	)	CT								
omiguralio	ni check - Neisseria r	UDIVIL	-21								
elper applicatio	ons										
Program	Path II	nstalled E	xecutal	le							
MBOSS infoalign	/usr/bin/infoalign	ok	ok								
EMBOSS sixpack	/usr/bin/sixpack	ok	ok								
MBOSS stretcher	/usr/bin/stretcher	ok	ok								
blastn	/usr/local/ncbi-blast+/bin/blastn	ok	ok								
blastp	/usr/local/ncbi-blast+/bin/blastp	ok	ok								
blastx	/usr/local/ncbi-blast+/bin/blastx	ok	ok								
ipcress	/usr/bin/ipcress	ok	ok								
mafft	/usr/local/bin/mafft	ok	ok								
makeblastdb /	usr/local/ncbi-blast+/bin/makeblastdb	ok	ok								
mogrify	/usr/bin/mogrify	ok	ok								
muscle	/usr/bin/muscle	ok	ok								
tblastx	/usr/local/ncbi-blast+/bin/tblastx	ok	ok								
ocus databases Locus	Database	Host	Port	Table	Primary id field	Secondary id field	Secondary id field valu	e Sequence field	Database accessible	Sequence quer	Sequences assigned
'16S rDNA	pubmlst_bigsdb_neisseria_seqdef	localhos	t 5432 :	sequences	allele_id	locus	'16S_rDNA	sequence	ok	ok	197
'porA	pubmlst_bigsdb_neisseria_seqdef	f localhos	t 5432 :	sequences	allele_id	locus	'porA	sequence	ok	ok	164
'porB	pubmlst_bigsdb_neisseria_seqdef	f localhos	t 5432	sequences	allele_id	locus	'porB	sequence	ok	ok	695
rpIF	pubmlst_bigsdb_neisseria_seqdef	f localhos	t 5432 :	sequences	allele_id	locus	rpIF	sequence	ok	ok	109
BACT000001 (rps/	<ul> <li>bigsdb_multispecies_seqdef</li> </ul>	localhos	t 5432	sequences	allele_id	locus	BACT000001	sequence	ok	ok	8189
BACT000002 (rpsE	<ol> <li>bigsdb_multispecies_seqdef</li> </ol>	localhos	t 5432	sequences	allele_id	locus	BACT000002	sequence	ok	ok	6567
BACT000003 (rpsC	C) bigsdb_multispecies_seqdef	localhos	t 5432	sequences	allele_id	locus	BACT000003	sequence	ok	ok	5965
BACT000004 (rpsE	D) bigsdb_multispecies_seqdef	localhos	t 5432 :	sequences	allele_id	locus	BACT000004	sequence	ok	ok	6195
BACT000005 (rpsE	<li>bigsdb_multispecies_seqdef</li>	localhos	t 5432	sequences	allele_id	locus	BACT000005	sequence	ok	ok	5707
		localhos	t 5432	sequences	allele_id	locus	BACT000006	sequence	ok	ok	4918
BACT000006 (rpsF											
BACT000006 (rpsF BACT000007 (rpsC			t 5432 :	sequences	allele_id	locus	BACT000007	sequence	ok	ok	5667

Any problems will be highlighted with a red X.

# 5.27 Exporting table configurations

Sometimes it is useful to transfer configurations between different databases or to export a configuration for troubleshooting. Data from most of the tables can be exported in tab-delimited text format suitable for batch uploading. For example, to export scheme configuration data, click the '?' link (Update or delete) next to schemes in the curator's interface.

3	tion			
Table	Add	Batch Add	Update or delete	Comments
loci	+	++	?	
	datab	ank scan		
locus aliases	+	++	?	Add alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci.
olate field extended attributes	+	++	?	Define additional attributes to associate with values of a particular isolate record field.
composite fields	+		?	Used to construct composite fields consisting of fields from isolate, loci or scheme fields.
sequence attributes	+	++		Define attributes that can be set for contigs in the sequence bin.
schemes	+	++	?	Describes schemes consisting of collections of loci, e.g. MLST.
scheme members	+	++	?	Defines which loci belong to a scheme.
scheme fields	+	++	?	Defines which fields belong to a scheme.
scheme groups	+	++	?	Describes groups in to which schemes can belong - groups can also belong to other groups.
scheme group scheme members	+	++	?	Defines which schemes belong to a group.
cheme group group members	+	++	?	Defines which scheme groups belong to a parent group.

Expand the filters and select the required scheme in the dropdown box, then press submit.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). @Log out   Change password	Toggle: 🚯
Query schemes for Neisseria PubMLST database	
Please enter your search criteria below (or leave blank and submit to return all records).  —Search criteria —Display—	
id • = • + 0 Order by: id • ascending •	
Display: 25 🗸 records per page 🖲	
Filter query by       Action         isolate display: <ul> <li>Reset</li> <li>Submit</li> <li>query field:</li> <li>query status:</li> <li>query status:</li></ul>	

Click the button 'Export configuration/data'.

Database: Species home   Curator's page (specie Users: Add   Query/update Isolates: Add   Query/update   Batch insert	es)   Curator's p	bage (database	e)		
Logged in: Keith Jolley (keith). HLog out   Change password					Toggle: 🕄
Query schemes for Neisseria PubMLS	T databa	ise			
Please enter your search criteria below (or leave blank and submit to re	eturn all record	s).			
Search criteria	C	Display			
id 🗸 = 🗸	+ 1	Order by: id		ascending	•
		Display: 25	5 👻 records per j	page 🕕	
─── ▼ Filter query by ─────	Action	۱ <u> </u>			
isolate display:	Reset	t Submit			
main display: 🗾 🕤 query field: 🔍 🗊					
query status:					
analysis:					
allow missing loci: 📃 🔽 🚺					
curator:	_				
scheme: MLST	• 0				
1 record returned.					
Delete Database configuration					
Delete ALL Export configuration/data					
	lbase dbase		isolate main		display allow
Delete Update (id/description) dbase name	host port	dbase table		query query field* status* <sup>6</sup>	analysis* order missing c
Delete Update 1 MLST pubmlst_bigsdb_neisseria_seqdef	n	nv_scheme_1	true true	true true	true 1
The second	reference.				4

The three tables that are used to define a scheme (schemes, scheme\_members and scheme\_fields) are displayed in a format suitable for copy and pasting.

```
schemes
____
id description dbase_name dbase_host dbase_port dbase_user dbase_password dbase_table isolate_dis
1 MLST pubmlst_bigsdb_neisseria_seqdef
                                               mv_scheme_1 1 1 1 1 1 1 2 2012-03-22
scheme_members
scheme_id locus profile_name field_order curator datestamp
1 abcZ 1 2 2009-11-12
1 adk
         2 2 2009-11-12
1 aroE
         3 2 2009-11-12
         4 2 2009-11-12
1 fumC
1 qdh
         5 2 2009-11-12
1 pdhC 6 2 2009-11-12
1 pgm 7 2 2009-11-12
scheme_fields
_____
scheme_id field type primary_key description field_order url isolate_display main_display
                                                                                         q
1 ST integer 1 1 /cgi-bin/bigsdb/bigsdb.pl?page=profileInfo&db=pubmlst_neisseria_seqdef&scheme
1 clonal_complex text 0 2 1 1 1 1 2 2009-11-16
```

# 5.28 Authorizing third-party client software to access authenticated resources

If you are running the *RESTful API*, you will need to specifically authorize client software to connect to authenticated resources. This involves creating a client key and a client secret that is used to sign requests coming from the application. The client key and secret should be provided to the application developer.

There is a script to do this in the scripts/maintenace directory of the download archive. The script is called create\_client\_credentials and should be run by the postgres user. A full list of options can be found by typing:

```
create_client_credentials.pl --help
NAME
   create_client_credentials.pl - Generate and populate
   authentication database with third party application (API client)
   credentials.
SYNOPSIS
   create_client_credentials.pl --application NAME [options]
OPTIONS
-a, --application NAME
   Name of application.
-d, --deny
   Set default permission to 'deny'. Permissions for access to specific
   database configurations will have to be set. If not included, the default
   permission will allow access to all resources by the client.
-h, --help
   This help page.
```

-i, --insert Add credentials to authentication database. This will fail if a matching application version already exists (use --update in this case to overwrite existing credentials).
-u, --update Update exisitng credentials in the authentication database.
-v, --version VERSION Version of application (optional).

# Curator's guide

Please note that links displayed within the curation interface will vary depending on database contents and the permissions of the curator.

### 6.1 Adding new sender details

All records within the databases are associated with a sender. Whenever somebody new submits data, they should be added to the users table so that their name appears in the dropdown lists on the data upload forms.

To add a user, click the add users (+) link on the curator's contents page.

Add, update or delete records           Record type         Add         Batch Add         Update or delete         Comments           users         +         +         ?         Users can be members of these groups - use for setting access permissions.           user groups         +         +         ?         Users can be members of these groups - use for setting access permissions.           user group members         +         +         ?         Add users for groups for setting access permissions.           user group members         +         +         ?         Add users or groups for setting access permissions.	
Record type         Add         Batch Add         Update or delete         Comments           USers         +         +         ?	
Record type         Add         Batch Add         Update or delete         Comments           USers         +         +         ?	_
Record type         Add         Batch Add         Update or delete         Comments           Users         +         +         ?	_
users         +         ++         ?           user groups         +         ++         ?         Users can be members of these groups - use for setting access permissions.           user group members         +         ++         ?         Add users to groups for setting access permissions.	_
users         +         ++         ?           user groups         +         ++         ?         Users can be members of these groups - use for setting access permissions.           user group members         +         ++         ?         Add users to groups for setting access permissions.	
User groups         +         +         ?         Users can be members of these groups - use for setting access permissions.           User group members         +         +         ?         Add users to groups for setting access permissions.	
user group members + ++ ? Add users to groups for setting access permissions.	
	of 'curator' in the users table.
projects + ++ ? Set up projects to which isolates can belong.	
project memories	
isolate allases + ++ ? Add alternative names for isolates.	
Sequences ++ ? The sequence bin holds sequence contigs from any source.	
accession number links + + + ? Associate sequences with Genbank/EMBL accession number.	
experiments + ++ ? Set up experiments to which sequences in the bin can belong.	
experiment sequences ? Add links associating sequences to experiments.	
experiment Sequences         ?         Add links associating sequences to experiments.           sequence tags         scan         ?         Tag regions of sequences within the sequence bin with locus information.	

Enter the user's details in to the form.

PrrhMILST Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert										
Logged in: Keith Jolley (keith). 🕪 Log out   Change password Toggle: 🕄										
Add new user	Add new user									
Record id:! user name:! surname:! first name:! email:!	Bloggs	Action Reset	Submit							
status:!										
date entered:	2015-06-18 2015-06-18									
	Keith Jolley (keith)									
	○ true ○ false 1 Receive new submission E-mails									

Normally the status should be set as 'user'. Only admins and curators with special permissions can create users with a status of curator or admin.

If the submission system is in operation there will be an option at the bottom called 'submission\_emails'. This is to enable users with a status of 'curator' or 'admin' to receive E-mails on receipt of new submissions. It is not relevant for users with a status of 'user' or 'submitter'.

## 6.2 Adding new allele sequence definitions

### 6.2.1 Single allele

To add a single new allele, click the sequences (all loci) add (+) link on the curator's main page - if only a few loci are defined with permission for the current user to curate then they will be listed individually and the specific locus allele addition links can also be used.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert										
Logged in: Keith Jolley (keith). Hog out   Change pass	ogged in: Keith Jolley (keith), BLog out   Change password									
Database curator's interfac	atabase curator's interface - Neisseria locus/sequence definitions									
Database curator s internat										
Add, update or delete records	Add	Batch Add	Update or delete	Comments						
users	+	++	?							
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.						
user group members	+	++	?	Add users to groups for setting access permissions.						
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.						
locus descriptions	+	++	?							
scheme curator control list	+	++	guery   batch	Define which curators can add or update profiles for particular schemes.						
locus curator control list	+	++	guery   batch	Define which curators can add or update sequences for particular loci.						
sequences (all loci)	+	++ I FASTA	?							
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.						
PubMed links (to sequences)	+	++	?							
Bexsero Antigen Sequence Typing (BAST) profiles	+	++	query   browse   list   batch update							
MLST profiles	+	++	query   browse   list   batch update							
rpIF species profiles	+	++	query   browse   list   batch update							
PubMed links (to profiles)	+	++								

Select the locus from the dropdown list box. The next available allele id will be entered automatically (if the allele id format is set to integer). Paste the sequence in to form, set the status and select the sender name from the dropdown box. If the sender does not appear in the box, you will need to add them to the registered users.

The status reflects the level of curation that the curator has done personally - the curator should not rely on assurances from the submitter. The status can either be:

- · Sanger trace checked
  - Sequence trace files have been assembled and inspected by the curator.
- WGS: manual extract (BIGSdb)
  - The sequence has been extracted manually from a BIGSdb database *by the curator*. There may be some manual intervention to identify the start and stop sites of the sequence.
- WGS: automated extract (BIGSdb)
  - The sequences have been generated by a BIGSdb tag scanning run and have had no manual inspection or intervention.
- WGS: visually checked
  - Short read data has been inspected visually using an alignment program by the curator.
- WGS: automatically checked
  - The sequences have been checked by an automated algorithm that assesses the quality of the data to ensure it meets specified criteria.
- unchecked
  - If none of the above match, then the sequence should be entered as unchecked.

	). @Log out   Change password	Help 🗹	Toggle:
d new allele	e sequence		
ise fill in the fields be	low - required fields are marked with an exclamation mark (!).		
Record			
locus:	abcZ		
allele id:!			
	TTTGATACCGTTGCCGAAGGTTTGGGCGAAATTCGCGATTTATTGCGCCGTTATCATCATG	TCACCOTC	
bequerioe	GCACAATTCAGACGACGTTTTATTAAAAGAACTCAACGAATTACAACTCGAAATCGAAGCG		
	TGGATGCGGCAGTCAAGCAGACTTTGGGCGAACTCGGTTTGCCGGAAAAACGAAAAAATCGG		
	CAGAAAAAGCGTGTCGCCTTGGCGCAGGCTTGGGTGCAGAAGCCCGACGTATTGCTGCTGG		
	TTTGGATATTGACGCGATTATCTGGTTGGAAAACCTGCTCAAGGCGTTTGAAGGCAGCTTG	GTCGTGATT	ACCCACO
	ATCGCCGTTTTTTGGACAATATCGCCACGCGGATTGTCGAACTCGATC		
-			.1
	Sanger trace checked 🗸		
	Jolley, Keith (keith)		
	Keith Jolley (keith)		
date entered:			
datestamp:!	2015-06-18		
comments:			
_			.:
Flags:			
	contains IS element		
	frameshift		
	internal stop codon 🔻 Use Ctrl click to select/deselect multiple choices		
PubMed ids:			
ENA ids:			
Genbank ids:	440		
Genbalik ida.			
<b>-</b>			
Override sequence			
Override sequence	e length check		
Action	_		
Reset Submit			

Press submit. By default, the system will test whether your sequence is similar enough to existing alleles defined for that locus. The sequence will be rejected if it isn't considered similar enough. This test can be overridden by checking the 'Override sequence similarity check' checkbox at the bottom. It will also check that the sequence length is within the allowed range for that locus. These checks can also be overridden by checking the 'Override sequence length check' checkbox, allowing the addition of unusual length alleles.

#### See also:

allele sequence flags

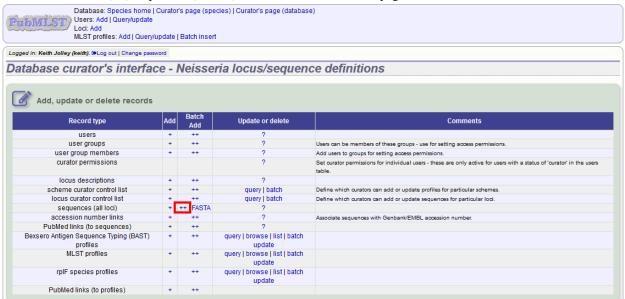
Sequences can also be associated with PubMed, ENA or Genbank id numbers by entering these as lists (one value per line) in the appropriate form box.

### 6.2.2 Batch adding multiple alleles

There are two methods of batch adding alleles. You can either upload a spreadsheet with all fields in tabular format, or you can upload a FASTA file provided all sequences are for the same locus and have the same status.

#### Upload using a spreadsheet

Click the batch add (++) sequences (all loci) link on the curator's main page.



Download a template Excel file from the following page.

PrrfrAILST Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Batch insert sequences	
<ul> <li>This page allows you to upload allele sequence data as tab-delimited text or copied from a spreadsheet.</li> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>If the locus uses integer allele ids you can leave the allele_id field blank and the next available number will be used.</li> <li>The status defines how the sequence was curated. Allowed values are: 'Sanger trace checked', 'WGS: manual extract', esquence flags can be added as a semi-colon (i) separated list</li> <li>Download tab-delimited header for your spreadsheet - use Paste special → text to paste the data.</li> <li>Download submission template (xisx format)</li> <li>Please note, some loci have extended attributes which may be required. For affected loci please use the batch insert preload page specific for locus: Select  <ul> <li>Reload</li> </ul> </li> <li>Please select the sender from the list below:</li> </ul>	
<ul> <li>✓ Ignore existing or duplicate sequences         <ul> <li>Ignore sequences containing non-nucleotide characters</li> <li>Silently reject all sequences that are not complete reading frames - these must have a start and in-frame stop cod sequences are also ignored.</li> <li>Override sequence similarity check</li> </ul> </li> <li>Paste in tab-delimited text (include a field header line).</li> </ul>	on at the ends and no internal stop codons. Existing Action Reset Submit

Fill in the spreadsheet. If the locus uses integer allele identifiers, the allele\_id can be left blank and the next available number will be used automatically. Paste the entire sheet in to the web form and select the sender from the dropdown box.

Additionally, there are a number of options available. Some of these will ignore sequences if they don't match certain criteria - this is useful when sequence data has been extracted from genomes automatically. Available options are:

- Ignore existing or duplicate sequences.
- Ignore sequences containing non-nucleotide characters.
- Silently reject all sequences that are not complete reading frames these must have a start and in-frame stop codon at the ends and no internal stop codons. Existing sequences are also ignored.
- Override sequence similarity check.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert			
Logged in: Keith Jolley (keith). Log out   Change password			Toggle: i
Batch insert sequences			
This page allows you to upload allele sequence data as tab-delimited text or copied from a spreadsheet. • Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish. • If the locus uses integer allele ids you can leave the allele_id field blank and the next available number will be used. • The status defines how the sequence was curated. Allowed values are: "Sanger trace checked", WGS: manual extract • Sequence flags can be added as a semi-colon (;) separated list • Download tab-delimited header for your spreadsheet - use Paste special> text to paste the data. • Download submission template (xisk format) • Please note, some loci have extended attributes which may be required. For affected loci please use the batch insert Reload page specific for locus: Select • Reload Please select the sender from the list below: Jolger existing or duplicate sequences Ignore existing or duplicate sequences Ignore sequences that are not complete reading frames - these must have a start and in-frame stop co sequences are also ignored. Override sequence similarity check Paste in tab-delimited text (include a field header line). Isocus allele_id sequence Status comments flags	page specific	to that locus:	

Press submit. You will be presented with a page indicating what data will be uploaded. This gives you a chance to back out of the upload. Click 'Import data'.

PubMLS	Database: Species home   Curator's page (species)   Curator's page (datab: Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	1SC)				
Logged in: Keith J	olley (keith). Log out   Change password					Toggle: i
Batch ins	ert sequences					
Import statu						
Sender: Keith Jo	lley					
No obvious prot	lems identified so far.					
Data to be in						
I ne following ta	ble shows your data. Any field with red text has a problem and needs to be checke	a. Note: Va	alla sequ	ience flags are	e displayed with a red b	ackground not red text.
locus allele_i					datestamp commen	its flags
abcZ 688	TTTGATACTGTTGCCGAAGG GCGGATTGTCGAACTCGATC Sanger trace checked		2		2014-07-09	
abcZ 689	TTTGATACTGTTGCCGAAGC GCGGATTGTCGAACTCGATC Sanger trace checked	d 2	2	2014-07-09	2014-07-09	

If there are any problems with the submission, these should be indicated at this stage, e.g.:

PubMLS	Database: Species home   Users: Add   Query/update Loci: Add MLST profiles: Add   Query/u		Curator's page (databas	e)					
Logged in: Keith Jo	olley (keith). Log out   Change passwo	rd							Toggle: i
Batch ins	ert sequences								
Import statu Primary k locus: abcZ; all		oblem(s) nucleotide (G A T C) charac	ters.						
Data to be in The following ta	, ble shows your data. Any field wi					_	displayed with a red b		not red text.
abcZ 688	TTTGATACTGTTGCCGAAGG			2	2	2014-07-09		nto nugo	
abcZ 689	TTTGATACTGTTGCCGAAGC	GCGGATTGTCGAACTCGATC	Sanger trace checked	2	2	2014-07-09	2014-07-09		

### Upload using a FASTA file

Uploading new alleles from a FASTA file is usually more straightforward than generating an Excel sheet.

Click 'FASTA' upload on the curator's contents page.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert									
gged in: Keith Jolley (keith). IBLog out   Change password									
Database curator's interface - Neisseria locus/sequence definitions									
	- I	10000	na locus/scquen						
Add, update or delete records Record type	Add	Batch Add	Update or delete	Comments					
users	+	++	?						
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.					
user group members	+	++	?	Add users to groups for setting access permissions.					
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.					
locus descriptions	+	++	?						
scheme curator control list	+	++	query   batch	Define which curators can add or update profiles for particular schemes.					
locus curator control list	+	++	query   batch	Define which curators can add or update sequences for particular loci.					
sequences (all loci)	+	++ FASTA	?						
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.					
PubMed links (to sequences)	+	++	?						
Bexsero Antigen Sequence Typing (BAST) profiles	+	++	query   browse   list   batch update						
MLST profiles	+	++	query   browse   list   batch update						
rpIF species profiles	+	++	query   browse   list   batch update						
PubMed links (to profiles)	+	++							

Select the locus, status and sender from the dropdown boxes and paste in the new sequences in FASTA format.

PubML	Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert
Logged in: Keith	Jolley (keith). Log out   Change password Toggle: 1
Batch in	sert sequences
(loci with integ	ows you to upload allele sequence data in FASTA format. The identifiers in the FASTA file will be used unless you select the option to use the next available id ger ids only). Do not include the locus name in the identifier in the FASTA file. nat you can not use this page to upload sequences for loci with extended attributes.
locus:	
	Sanger trace checked 🗸
sender:!	Jolley, Keith (keith)
(FASTA):!	>isolate1 ITTGATACTGTTGCCGAAGGTTTGGGCGAAATTCGCGATTTATTGCAACGTTATCATCAT GTCAGCCATGGAGTCGGAAGGACGGCCGAAGGGCGGCGGCGGCGGCGGCGGCGG
sequences	all sequences that are not complete reading frames - these must have a start and in-frame stop codon at the ends and no internal stop codons. Existing are also ignored. e sequence similarity check xt available id (only for loci with integer ids) 

For loci with integer ids, the next available id number will be used by default (and the identifier in the FASTA file will be ignored). Alternatively, you can indicate the allele identifier within the FASTA file (do not include the locus name in this identifier).

As with the spreadsheet upload, you can select options to ignore selected sequences if they don't match specific criteria.

#### Click 'Check'.

The sequences will be checked. You will be presented with a page indicating what data will be uploaded. This gives you a chance to back out of the upload. Click 'Upload valid sequences'.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert								
Logged in: Keith Jolley (keith). Log out   Change password								
Batch insert sequences	Batch insert sequences							
Sequence check Original designation Allele id Status isolate1 688 OK isolate2 689 OK	Action Upload valid sequences							

Any invalid sequences will be indicated in this confirmation page and these will not be uploaded (you can still upload the others), e.g.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert								
Logged in: Keith Jolley (	ogged in: Keith Jolley (keith). Log out   Change password Toggle:							
Batch insert	Batch insert sequences							
Sequence check Original designation isolate1 isolate2		I Status OK Sequence contains non nucleotide (G[A[T]C) characters.	Action Upload valid sequences					

# 6.3 Updating and deleting allele sequence definitions

Note: You cannot update the sequence of an allele definition. This is for reasons of data integrity since an allele may

form part of a scheme profile and be referred to in multiple databases. If you really need to change a sequence, you will have to remove the allele definition and then re-add it. If the allele is a member of a scheme profile, you will also have to remove that profile first, then re-create it after deleting and re-adding the allele.

In order to update or delete an allele, first you must select it. Click the query (?) sequences (all loci) link - if only a few loci are defined with permission for the current user to curate then they will be listed individually and the specific locus query links can also be used.

Database: Species home   Users: Add   Querylupdate Loci: Add MLST profiles: Add   Querylu			cies)   Curator's page (databas	9)
Logged in: Keith Jolley (keith). Hog out   Change passw	vord			
Database curator's interfac	:e -	Neisse	ria locus/sequen	ce definitions
Add, update or delete records	Add	Batch	Update or delete	Comments
USERS	+	Add	2	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	2	Add users to groups for setting access permissions.
curator permissions			?	Set curstor by groups to standy docump permission. Set curstor permissions for individual users - these are only active for users with a status of 'curator' in the users table.
locus descriptions	+	++	?	
scheme curator control list	+	++	query   batch	Define which curators can add or update profiles for particular schemes.
locus curator control list	+	++	query   batch	Define which curators can add or update sequences for particular loci.
sequences (all loci)	+	++   FASTA	?	
accession number links	+	++	2	Associate sequences with Genbank/EMBL accession number.
PubMed links (to sequences)	+	++	2	
Bexsero Antigen Sequence Typing (BAST)	+	++	query   browse   list   batch	
profiles			update	
MLST profiles	+	++	query   browse   list   batch update	
rpIF species profiles	+	++	query   browse   list   batch update	
PubMed links (to profiles)	+	++		

Either search for specific attributes in the search form, or leave it blank and click 'Submit' to return all alleles. For a specific allele, select the locus in the filter and enter the allele number in the allele\_id field.

PubMLST Users:	Add   Query/update	or's page (species)   Curato e   Batch insert	r's page (databa	ise)	
Logged in: Keith Jolley (keith). Lo	og out   Change password				Toggle: i
Query sequence	s for Neisseri	a locus/sequei	nce defin	itions database	
Some loci have additional fie export plugins.	elds which are not searcha	able from this general page	. Search for thes	e at the locus-specific query page. Us	e this page also for access to the sequence analysis or
Also note that some loci in the to a locus that uses integer a			ueries using the	'<' or '>' modifiers will work alphabetic	ally rather than numerically unless you filter your search
Please enter your search cri	teria below (or leave blank	and submit to return all rec	cords).		
			Display		
allele id 👻 =	= <b>v</b> 4	+ <u>i</u>	Order by:	locus 👻 ascending	
			Display:	25 👻 records per page i	
		Action			
locus:	abcZ	▼ i Reset	Submit		
status:		•			
sender:		▼ i			
curator:		i			
allele flag:	<b></b>				

Click the appropriate link to either update the allele attributes or to delete it. If you have appropriate permissions, there may also be a link to 'Delete ALL'. This allows you to quickly delete all alleles returned from a search.

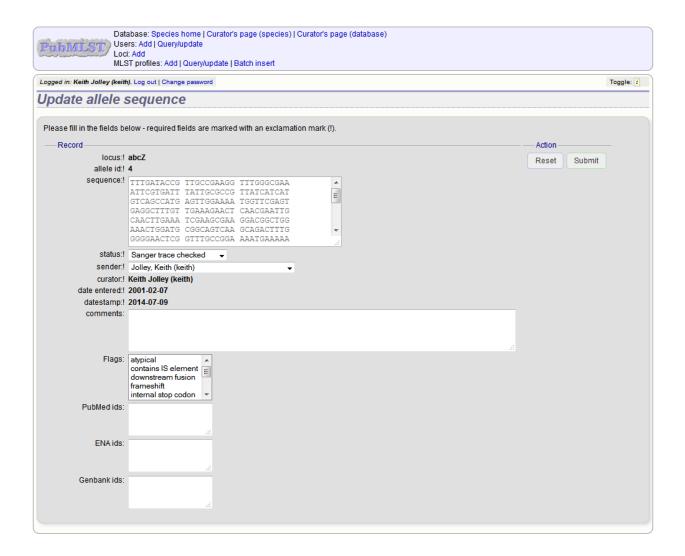
Database: Species home   Curator's page (species)   Curator's page (database)         Users: Add   Query/update         MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	oggle: i
Query sequences for Neisseria locus/sequence definitions database	
Some loci have additional fields which are not searchable from this general page. Search for these at the locus-specific query page. Use this page also for access to the sequence analysis or export page.	olugins.
Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a loc uses integer allele ids using the drop-down list.	us that
Please enter your search criteria below (or leave blank and submit to return all records).	
Search criteria	
allele id • = • 4 + [] Order by: locus • ascending • Display: 25 • records per page []	
Filter query by     Action       locus:     abcZ       status:     v       sender:     v       curator:     v       allele flag:     v	
1 record returned.	
Delete ALL Export configuration/data Batch set	
Delete  Update   locus   allele id         sequence         sequence length         status         sender         curator         date entered         datestamp   comments         flags           Delete Update         abcZ         4         TITGALACCETTECC         TITGALACCETTECC         433         Sanger trace checked         Keith Jolley Man-Suen Chan         2001-02-07         2009-11-11	

If you choose to delete, you will be presented with a final confirmation screen. To go ahead, click 'Delete!'. Deletion will not be possible if the allele is part of a scheme profile - if it is you will need to delete any profiles that it is a member of first.

PubMLST L	vatabase: Species home   Curator's page (species)   Curator's page (database) Sers: Add   Query/update oci: Add ILST profiles: Add   Query/update   Batch insert
Logged in: Keith Jolley (ke	eith). Log out   Change password
Delete allele	sequence
You have chosen to de	elete the following record:
locus:	abcZ
allele id:	4
sequence:	ITIGATACCG ITGCCGAAGG ITIGGGCGAA AITCGIGAIT IAITGCGCCG ITAICAICAI GICAGCCAIG AGIIGGAAAA IGGIICGAGI GAGGCIIIGI IGAAGAACI CAACGAAIIG CAACIGAA ICGAAGCGAA GGACGGCIGG AAACIGGAIG CGGCAGICA GCAGCAIIG GGGGAACICG GIIGGCGGA AAAIGAAAA AICGGCAACC ITICGGGGG ICAGAAAAA CGGCIGCCI IGGCIACGA IIGGGIGCAA AAGCCCGACG IAITGCIGCIG GACGAGCCG ACCAACCAII IGGAIAICGA CGCGAIIAII IGGCIGGAAA AICIGCICAA AGCGIIIGAA GGCAGCIIG IIGIGAIIAC CCACGACCG CIIIIIIGG ACAAIAICCC CACGGGGAI GICGAACIGG AIC
status:	Sanger trace checked
sender:	Keith Jolley
curator:	Man-Suen Chan
date entered:	2001-02-07
datestamp:	2009-11-11
comments:	
Delete!	

If instead you clicked 'Update', you will be able to modify attributes of the sequence, or link PubMed, ENA or Genbank records to it. You will not be able to modify the sequence itself.

Note: Adding flags and comments to an allele record requires that this feature is enabled in the *database configuration*.



## 6.4 Retiring allele identifiers

Sometimes there is a requirement to prevent the automated assignment of a particular allele identifier - an allele with that identifier may have been commonly used and has since been removed. Reassignment of the identifier to a new sequence may lead to confusion, so in this instance, it would be better to prevent this.

You can retire an allele identifier by clicking the 'Add' retired allele ids link on the sequence database curators' page.

ProfotATLST Users: A Loci: gtr	dd   0   mut	Query/update S   yqiL   gki	Curator's page (species)   C murl   recP   xpt /update   Batch insert	turator's page (database)
Logged in: Keith Jolley (keith). 🗭	.og ou	t   Change pass	sword	
Database curato	r's	interfa	ce - Streptococ	cus pyogenes locus/sequence definitions
Add, update or d	lelet			
Record type	Add	Batch Add	Update or delete	Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++		Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
locus descriptions	+	++	?	
scheme curator control list	+	++	query   batch	Define which curators can add or update profiles for particular schemes.
locus curator control list	+	++	query   batch	Define which curators can add or update sequences for particular loci.
sequences (all loci)	+	++   FASTA	?	
gki sequences	+	++	?	
gtr sequences	+	++	?	
murl sequences	+	++	?	
mutS sequences	+	++	?	
recP sequences	+	++	?	
xpt sequences	+	++	?	
yqiL sequences	+	++	?	
retired allele ids	+	++	?	Allele ids defined here will be prevented from being used.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
PubMed links (to sequences)	+	++	?	
MLST profiles	+	++	query/browse/list   batch update	
PubMed links (to profiles)	+	++		

Select the locus from the dropdown list box and enter the allele id. Click 'Submit'.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loc: gtr   mutS   yqiL   gki   murl   recP   xpt MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). (+Log out   Change password	Toggle: 🚯
Add new retired allele id	
Please fill in the fields below - required fields are marked with an exclamation mark (!).   Record  Iocus:1 yqL  Action  Reset Submit  curator! Keith Jolley (keith)	
datestamp:1 2015-10-19	

You cannot retire an allele that already exists, so you must delete it before retiring it. Once an identifier is retired, you will not be able to create a new allele with that name.

# 6.5 Updating locus descriptions

Loci in the sequence definitions database can have a description associated with them. This may contain information about the gene product, the biochemical reaction it catalyzes, or publications providing more detailed information etc. This description is accessible from various pages within the interface such as an *allele information page* or from the *allele download page*.

**Note:** In recent versions of BIGSdb, a blank description record is created when a new locus is defined. The following instructions assume that this is the case. It is possible for this record to be deleted or it may never have existed if the locus was created using an old version of BIGSdb. If the record does not exist, it can be added by clicking the Add (+) button next to 'locus descriptions'. Fill in the fields in the same way as described below.

To edit a locus description, first you need to find it. Click the update/delete (?) button next to 'locus descriptions' on the sequence database curator's page (depending on the permissions set for your user account not all the links shown here may be displayed).

Users: Add   Query/update Loci: Add MLST profiles: Add   Query/u	ipdate	Batch inser	t	
gged in: Keith Jolley (keith). 🗭 Log out   Change passv	vord			
atabase curator's interfac	:e -	Neisse	ria locus/sequen	ce definitions
Add, update or delete records	Add	Batch Add	Update or delete	Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
locus descriptions	+	++	?	
scheme curator control list	+	++	guery   batch	Define which curators can add or update profiles for particular schemes.
locus curator control list	+	++	guery   batch	Define which curators can add or update sequences for particular loci.
sequences (all loci)	+	++   FASTA	?	
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
PubMed links (to sequences)	+	++	?	
Bexsero Antigen Sequence Typing (BAST) profiles	+	++	query   browse   list   batch update	
MLST profiles	+	++	query   browse   list   batch update	
	+	++	query   browse   list   batch	
rpIF species profiles			update	

Either enter the name of the locus in the query box:

Database: Species home   Curator's page (species)   Curator's Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	page (database)
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: 1
Query locus descriptions for Neisseria locus/	sequence definitions database
Please enter your search criteria below (or leave blank and submit to return all recor — Search criteria —	ds). Display —————
locus 🗸 = 🗸 NEIS0620 + i	Order by: locus 👻 ascending 👻
	Display: 25 🗸 records per page i
— ▷ Filter query by — Action     Reset     Submit	

or expand the filter list and select it from the dropdown box:

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Query locus descriptions for Neisseria locus/sequence definitions database	
Please enter your search criteria below (or leave blank and submit to return all records).  Search criteria  Iocus	
Action Action	
curator:	
common name:	

Click 'Submit'.

If the locus description exists, click the 'Update' link (if it doesn't, see the note above).

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Query locus descriptions for Neisseria locus/sequence definitions database	
Please enter your search criteria below (or leave blank and submit to return all records).  —Search criteria	
locus 🗸 = 🗸 + i Order by: locus 🗸 ascending 🗸	
Display: 25	
1 record returned.	
Delete — Database configuration — Delete ALL Export configuration/data	
Delete         Update         locus         full name         product         description         curator         datestamp           Delete         Update         VEIS0620         malate oxidoreductase (EC 1.1.1.38)         Final step in TCA cycle producing oxaloacetate.         Keith Jolley         2010-10-28	

#### Fill in the form as needed:

PribMLST Users	asse: Species home   Curator's page (species)   Curator's page (database) : Add   Queny/update Add   profiles: Add   Queny/update   Batch insert	
Logged in: Keith Jolley (keith).	Log out   Change password	Toggle: $[i]$
Update locus d	lescription	
Please fill in the fields belo	w - required fields are marked with an exclamation mark (!).	
	NEIS0620 Keith Jolley (keith) 2014-11-04	Action Reset Submit
full name:		
product	.ii malate oxidoreductase (EC 1.1.1.38)	
description	ł.	
	Final step in TCA cycle producing oxaloacetate.	
aliases:	NMA0870 NMB0671 NMC0620	
PubMed ids:	14917678 	
links: (Format: URL description)	http://www.enzyme-database.org /query.php?ec=1.1.1.38/EC 1.1.1.38	

• full\_name

The full name of the locus - often this can be left blank as it may be the same as the locus name. An example of where it is appropriately used is where the locus name is an abbreviation, e.g. PorA\_VR1 - here we could enter 'PorA variable region 1'. This should not be used for the 'common name' of the locus (which is defined within the locus record itself) or the gene product.

• product

The name of the protein product of a coding sequence locus.

· description

This can be as full a description as possible. It can include the specific part of the biochemical pathway the gene product catalyses or may provide background information, as appropriate.

aliases

These are alternative names for the locus as perhaps found in different genome annotations. Don't duplicate the locus name or common name defined in the locus record. Enter each alias on a separate line.

• Pubmed\_ids

Enter the PubMed id of any paper that specifically describes the locus. Enter each id on a separate line. The software will retrieve the full citation from PubMed (this happens periodically so it may not be available for display immediately).

• Links

Enter links to additional web-based resources. Enter the URL first followed by a pipe symbol (I) and then the description.

Click 'Submit' when finished.

### 6.6 Adding new scheme profile definitions

Provided a scheme has been set up with at least one locus and a scheme field set as a primary key, there will be links on the curator's main page to add profiles for that scheme.

To add a single profile you can click the add (+) profiles link next to the scheme name (e.g. MLST):

MLST profiles: Add   Query/u		Batch Inser	t	
ed in: Keith Jolley (keith). DLog out   Change pass	word			
tabase curator's interfac	ce - )	Neisse	ria locus/sequen	ce definitions
~				
Add, update or delete records				
Descent twee		Batch	Undete en delete	0 million to
Record type	Add	Add	Update or delete	Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
locus descriptions	+	++	?	
scheme curator control list	+	++	query   batch	Define which curators can add or update profiles for particular schemes.
locus curator control list	+	++	query   batch	Define which curators can add or update sequences for particular loci.
sequences (all loci)	+	++   FASTA	?	
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
PubMed links (to sequences)	+	++	?	
Bexsero Antigen Sequence Typing (BAST)	+	++	query   browse   list   batch update	
profiles		++	query   browse   list   batch	
profiles MLST profiles			update	
	+	**	update query   browse   list   batch update	

A form will be displayed with the next available primary key number already entered (provided integers are used for the primary key format). Enter the new profile, associated scheme fields, and the sender, then click 'Submit'. The new profile will be added provided the primary key or the profile has not previously been entered.

PubMLST)	Database: Species home   Curator's page (species)   Curator's page (database) Jsers: Add   Query/update .oci: Add /LST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (ke	reith). Log out   Change password Togg	le: i
Add new ML	ST profile	
	below - required fields are marked with an exclamation mark (!).	
Record		
	10880	
abcZ: !		
adk: !		
aroE: ! fumC: !		
gdh: !		
pdhC: !		
pano. I pgm: !		
sender: !		
clonal_complex:		
	Keith Jolley (keith)	
date_entered: !		
datestamp: !		
PubMed ids:		
Action		
Reset Subn	mt	

More usually, profiles are added in a batch mode. It is often easier to do this even for a single profile since it allows copying and pasting data from a spreadsheet.

Click the batch add (++) profiles link next to the scheme name:

Database: Species home   Curator's page (species)   Curator's page (database) Prr 17/1/J.CT Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert								
Logged in: Keith Jolley (keith). HLog out   Change passw	ord							
Database curator's interfac	e - I	Neisse	ria locus/sequen	ce definitions				
Add, update or delete records		Batch						
Record type	Add	Add	Update or delete	Comments				
users	+	++	?					
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.				
user group members	+	++	?	Add users to groups for setting access permissions.				
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.				
locus descriptions	+	++	?					
scheme curator control list	+	++	query   batch	Define which curators can add or update profiles for particular schemes.				
locus curator control list	+	++	query   batch	Define which curators can add or update sequences for particular loci.				
sequences (all loci)	+	++   FASTA	?					
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.				
PubMed links (to sequences)	+	++	?					
Bexsero Antigen Sequence Typing (BAST) profiles	+	++	query   browse   list   batch update					
MLST profiles	+	++	query   browse   list   batch update					
rpIF species profiles	+	++	query   browse   list   batch update					
PubMed links (to profiles)	+	++						

Click the 'Download submission template (xlsx format)' link to download an Excel submission template.

PTIDIMLST PTIDIMLST Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Batch insert MLST profiles	
<ul> <li>This page allows you to upload profiles as tab-delimited text or copied from a spreadsheet.</li> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>You can choose whether or not to include a ST field - if it is omitted, the next available ST will be used automatically. also provide it for each profile record.</li> <li>Download tab-delimited header for your spreadsheet - use Paste special → text to paste the data.</li> <li>Download submission template (xisx format)</li> <li>Please paste in tab-delimited text (include a field header line)</li> </ul>	If however, you include it in the header line, then you must

Fill in the spreadsheet using the copied template, then copy and paste the whole spreadsheet in to the large form on the upload page. If the primary key has an integer format, you can exclude this column and the next available number will be used automatically. If the column is included, however, a value must be set. Select the sender from the dropdown list box and then click 'Submit'.

Database: Species home   Curator's page (species home   Curato	ies)   Curator's page (database)	
Logged in: Keith Jolley (keith). Log out   Change password		Toggle: i
Batch insert MLST profiles		
This page allows you to upload profiles as tab-delimited text or copied		
<ul> <li>Field header names must be included and fields can be in an</li> <li>You can choose whether or not to include a ST field - if it is om you must also provide it for each profile record.</li> </ul>		however, you include it in the header line, then
Download tab-delimited header for your spreadsheet - use Pa	ste special $\rightarrow$ text to paste the data.	
Download submission template (xlsx format)		
Please paste in tab-delimited text (include a field header line) abcz adk arot fumc gdh pdhc	pgm clonal_complex	_
	6	
Parameters	-Action	
Sender: Jolley, Keith (keith)	Reset Submit	
Value will be overridden if you include a sender field in your pasted data.		
Ignore duplicate profiles     Back		

You will be given a final confirmation page stating what will be uploaded. If you wish to proceed with the submission, click 'Import data'.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Batch insert MLST profiles	
Import status	
Sender: Keith Jolley	
No obvious problems identified so far.	
Import data	
Data to be imported	
The following table shows your data. Any field coloured red has a problem and needs to be checked.	
ST     abcZ     adk     aroE     fumC     gdh     pdhC     pgm     clonal_complex     sender     curator     date_entered     datestamp       10880     2     3     4     122     8     4     6     2     2     2014-07-10     2014-07-10	

# 6.7 Updating and deleting scheme profile definitions

In order to update or delete a scheme profile, first you must select it. Click the query (?) profiles link next to the scheme name (e.g. MLST):

PrrhMLST Users: Add   Query/update Loc: Add MLST profiles: Add   Query/update			ecies)   Curator's page (databas t	e)
Logged in: Keith Jolley (keith). Dog out   Change pass	vord			
Database curator's interfac	:e -	Neisse	ria locus/sequen	ce definitions
Add, update or delete records		Detab	•	
Record type	Add	Batch Add	Update or delete	Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
locus descriptions	+	++	?	
scheme curator control list	+	++	query   batch	Define which curators can add or update profiles for particular schemes.
locus curator control list	+	++	query   batch	Define which curators can add or update sequences for particular loci.
sequences (all loci)	+	++   FASTA	?	
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
PubMed links (to sequences)	+	++	?	
Bexsero Antigen Sequence Typing (BAST) profiles	+	++	query   browse   list   batch update	
MLST profiles	+	++	query browse   list   batch update	
rpIF species profiles	+	++	query   browse   list   batch update	
PubMed links (to profiles)	+	++		

Search for your profile by entering search criteria (alternatively you can use the browse or list query functions).

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Query/update profiles - Neisseria locus/sequence definitions	
Schemes Please select the scheme you would like to query: MLST  Select	
Locus/scheme fields     Display/sort options       ST     • = • 4563       T     • I       Order by:     ST       Display:     25 • records per page []	
1 record returned.         Delete         Delete ALL         Delete IUpdate       ST         abcZ       adk         aroE       fumC         gdh       pdhC         pgm       clonal complex         Delete       Update         4563       2       7       6       13       9       18       8       ST-167 complex	

To delete the profile, click the 'Delete' link next to the profile. Alternatively, if your account has permission, you may be able to 'Delete ALL' records retrieved from the search.

For deletion of a single record, the full record will be displayed. Confirm deletion by clicking 'Delete!'.

PubMLST)	Loci: Add MLST profiles: Add   Query/update   Batch insert								
Logged in: Keith Jolley (ke	Log out   Change password								
Delete profile	2								
You have chosen to de	elete the following record:								
scheme id:	1) MLST								
ST:	4563								
abcZ:	2								
adk:	7								
aroE :	6								
fumC :	13								
gdh :	9								
pdhC :	18								
pgm :	8								
clonal_complex :	ST-167 complex								
sender:	Ana-Belen Ibarz-Pavon								
curator:	Keith Jolley								
date entered:	2005-03-03								
datestamp:	2009-11-11								
Delete!									

To modify the profile, click the 'Update' link next to the profile following the query. A form will be displayed - make any changes and then click 'Update'.

PubMLST)	vatabase: Species home   Curator's page (species)   Curator's page (da Isers: Add   Query/update oci: Add ILST profiles: Add   Query/update   Batch insert	tabase)	
Logged in: Keith Jolley (ke	eith). Log out   Change password		Toggle: i
Update profil	le		
Record		Action	
Update your record a	as required - required fields are marked with an exclamation mark (!):	Reset Submit	
ST: !	4563		
abcZ: !	2		
adk: !			
aroE: !	6		
fumC: !			
gdh: !			
pdhC: !	18		
pgm: !			
	ST-167 complex		
	Ibarz-Pavon, Ana-Belen (aibarz) 🗸 🗸		
	Keith Jolley (keith)		
date_entered: !			
datestamp: ! PubMed ids:			
Pubmedilds.			

# 6.8 Adding isolate records

To add a single record, click the add (+) isolates link on the curator's index page.

Add, update or delete record Record type Add Ba Users + user groups + user groups +	atch Add Update	e or delete	Comments
Record type         Add         Ba           users         +           user groups         +	atch Add Update		Comments
Record type         Add         Ba           users         +           user groups         +	atch Add Update		Comments
users + user groups +	++		Comments
user groups +			
		?	
user group members +	++	? Use	ers can be members of these groups - use for setting access permissions.
	++	? Add	d users to groups for setting access permissions.
curator permissions		? Set	t curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
isolates +	++ query   browse	list   batch update	
olate field extended attribute values +	++	? Add	d values for additional isolate field attributes.
projects +	++	? Set	t up projects to which isolates can belong.
project members +	++	? Add	d isolates to projects.
isolate aliases +	++	? Add	d alternative names for isolates.
PubMed links +	++	?	
allele designations	++	? Alle	ele designations can be set within the isolate table functions.
sequences	++	? The	e sequence bin holds sequence contigs from any source.
accession number links +	++	? Ass	sociate sequences with Genbank/EMBL accession number.
experiments +	++	? Set	t up experiments to which sequences in the bin can belong.
experiment sequences		? Ado	d links associating sequences to experiments.
sequence tags so	can	? Tag	g regions of sequences within the sequence bin with locus information.

The next available id will be filled in automatically but you are free to change this. Fill in the individual fields. Required fields are listed first and are marked with an exclamation mark (!). Some fields may have drop-down list boxes of allowed values. You can also enter allele designations for any loci that have been defined.

PubMLST / Users: A	se: Species home   Curator's ; Add   Query/update :: Add   Query/update   Batch in		s)   Curator's page	e (database)					
Logged in: Keith Jolley (keith). Lo	og out   Change password								Toggle: i
Add new isolate									
Please fill in the fields below	- required fields are marked v	vith an exclan	nation mark (!).						
Isolate fields					Allele design	nations			
id:!	30457 🔷 [	i							
isolate:!	J323_2 i						MLST		
country:		i			abcZ	adk	aroE	fumC	
species:!	Neisseria meningitidis		<b>▼</b> []						
	Jolley, Keith (keith)		▼ 1		gdh	pdhC	pgm		
	Keith Jolley (keith)								
	2014-07-10 👔					Fin	etyping antig	ens	
datestamp:! region:	2014-07-10 👔				PorA VR1	PorA VR2	FetA VR		
-	2014	i			POTA VKT	POTA VKZ	FELAVK		
epidemiological year:		1			A				
age yr:		i			-Action				
age mth:					Reset S	ubmit			
Sex:									
disease:	meningitis	<b>√</b> i							
source:	CSF v i								
epidemiology:									
serogroup:	B 👻 👔								
MLEE designation:	<b>_</b>								
serotype:									
sero subtype:									
ET no:									
penicillin:	i								
penicillin range:		i							
amoxicillin:	i								
sulphonamide:	i								
ceftriaxone:	i								

Press submit when finished. More usually, isolate records are added in batch mode, even when only a single record is added, since the submission can be prepared in a spreadsheet and copied and pasted.

Select batch add (++) isolates link on the curator's index page.

isolates     +     +       plate field extended attribute values     +     +       projects     +     +       project members     +     +       isolate aliases     +     +       extreme     +     +	be members of these groups - use for setting access permissions. to groups for setting access permissions. r permissions for individual users - these are only active for users with a status of 'curator' in the users table. is for additional isolate field attributes. jects to which isolates can belong. tes to projects.	4 - k	snang	e password			
Add, update or delete records         Record type       Add       Batch Add       Update or delete       Comments         users       +       ++       ?       Users can be members of these groups - use for setting access permissions.         user groups       +       ++       ?       Add users to groups for setting access permissions.         user group members       +       +       ?       Add users to groups for setting access permissions.         curator permissions       ?       Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user isolates         late field extended attribute values       ++       ?       Add values for additional isolate field attributes.         projects       ++       ?       Add values for additional isolates an belong.         projects       ++       ?       Add isolates to projects.         isolate aliases       +       ?       Add alternative names for isolates.	be members of these groups - use for setting access permissions. to groups for setting access permissions. r permissions for individual users - these are only active for users with a status of 'curator' in the users table. is for additional isolate field attributes. jects to which isolates can belong. tes to projects.		-	ufo o o	Najaaari		۲.
Record type         Add         Batch Add         Update or delete         Comments           users         +         ++         ?         Users can be members of these groups - use for setting access permissions.           user group members         +         +         ?         Add users to groups for setting access permissions.           user group members         +         +         ?         Add users to groups for setting access permissions.           curator permissions         ?         Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user           isolates         +         +         ?         Add values for additional isolate field attributes.           projects         +         +         ?         Add values for additional isolates an belong.           projects         +         +         ?         Add isolates and points.           isolate aliases         +         +         ?         Add isolates to projects.	be members of these groups - use for setting access permissions. to groups for setting access permissions. r permissions for individual users - these are only active for users with a status of 'curator' in the users table. is for additional isolate field attributes. jects to which isolates can belong. tes to projects.	tabase curator's in	ne	rrace	- Neisseri	a Pudmls	
Record type         Add         Batch Add         Update or delete         Comments           users         +         ++         ?         Users can be members of these groups - use for setting access permissions.           user group members         +         +         ?         Add users to groups for setting access permissions.           user group members         +         +         ?         Add users to groups for setting access permissions.           curator permissions         ?         Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user isolates           isolates         +         +         ?         Add values for additional isolate field attributes.           projects         +         +         ?         Set up rojects to which isolates an belong.           projects         +         +         ?         Add values for additional isolates field attributes.           isolate aliases         +         +         ?         Add selates to projects.	be members of these groups - use for setting access permissions. to groups for setting access permissions. r permissions for individual users - these are only active for users with a status of 'curator' in the users table. is for additional isolate field attributes. jects to which isolates can belong. tes to projects.						
Record type         Add         Batch Add         Update or delete         Comments           users         +         ++         ?         Users can be members of these groups - use for setting access permissions.           user group members         +         +         ?         Add users to groups for setting access permissions.           user group members         +         +         ?         Add users to groups for setting access permissions.           curator permissions         ?         Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user isolates           isolates         +         +         ?         Add values for additional isolate field attributes.           projects         +         +         ?         Set up rojects to which isolates an belong.           projects         +         +         ?         Add values for additional isolates field attributes.           isolate aliases         +         +         ?         Add selates to projects.	be members of these groups - use for setting access permissions. to groups for setting access permissions. r permissions for individual users - these are only active for users with a status of 'curator' in the users table. is for additional isolate field attributes. jects to which isolates can belong. tes to projects.						
users     +     ++     ?       user groups     +     ++     ?     Users can be members of these groups - use for setting access permissions.       user group members     +     ++     ?     Add users to groups for setting access permissions.       user group permissions     -     ?     Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user isolates       user field extended attribute values     +     ++     ?     Add values for additional isolate field attributes.       projects     +     ++     ?     Add values for additional isolate field attributes.       projects     +     ++     ?     Add values for additional isolate field attributes.       projects     +     ++     ?     Add values for additional isolate field attributes.       isolate aliases     +     ++     ?     Add values for projects.	be members of these groups - use for setting access permissions. to groups for setting access permissions. r permissions for individual users - these are only active for users with a status of 'curator' in the users table. is for additional isolate field attributes. jects to which isolates can belong. tes to projects.	Add, update or delete	reco	ords			
users     +     ++     ?       user groups     +     ++     ?     Users can be members of these groups - use for setting access permissions.       user group members     +     ++     ?     Add users to groups for setting access permissions.       curator permissions     -     ?     Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user isolates       isolates     +     ++     ?     Add values for additional isolate field attributes.       projects     +     ++     ?     Add values for additional isolate field attributes.       projects     +     ++     ?     Add values for additional isolate field attributes.       projects     +     ++     ?     Add values for additional isolate field attributes.       isolate aliases     +     +     ?     Add values for projects.	to groups for setting access permissions. or permissions for individual users - these are only active for users with a status of 'ourator' in the users table. es for additional isolate field attributes. a before to which isolates can belong. tes to projects.	Record type	Add	Batch Ad	ld Update	e or delete	Comments
user group members       +       +++       ?       Add users to groups for setting access permissions.         curator permissions       ?       Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user'         isolates       +++       ?       Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user'         projects       +       ++       ?       Add values for additional isolate field attributes.         projects       +       ++       ?       Add values for additional isolate field attributes.         projects       +       +       ?       Add values to projects to which isolates can belong.         isolate aliases       +       +       ?       Add isolates to projects.	to groups for setting access permissions. or permissions for individual users - these are only active for users with a status of 'ourator' in the users table. es for additional isolate field attributes. a before to which isolates can belong. tes to projects.		+				
user group members         +         ++         ?         Add users to groups for setting access permissions.           curator permissions         ?         Seturator permissions for individual users - these are only active for users with a status of 'curator' in the user           isolates         +         ++         ?         Seturator permissions for individual users - these are only active for users with a status of 'curator' in the user           projects         +         ++         ?         Add values for additional isolate field attributes.           projects         +         ++         ?         Set up rojects to which isolates can belong.           projects         +         +         ?         Add values to projects.           isolate allases         +         +         ?         Add alternative names for isolates.	or permissions for individual users - these are only active for users with a status of 'curator' in the users table. It is for additional isolate field attributes. Jjects to which isolates can belong. Les to projects.	user groups	+	++		?	Users can be members of these groups - use for setting access permissions.
isolates     +++     query   browse   list   batch update       late field extended attribute values     +     +       projects     +     +       project members     +     +       isolate aliases     +     +       ?     Add values for additional isolate field attributes.       biolate aliases     +     +	s for additional isolate field attributes.		+	++		?	Add users to groups for setting access permissions.
Idlet field extended attribute values     +     +     ?     Add values for additional isolate field attributes.       projects     +     +     ?     Set up projects to which isolates can belong.       project members     +     +     ?     Add isolates to projects.       Isolate allases     +     +     ?     Add attermative names for isolates.	ojects to which isolates can belong. Les to projects.	curator permissions				?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
projects     +     +++     ?     Set up projects to which isolates can belong.       project members     +     +++     ?     Add isolates to projects.       isolate aliases     +     +++     ?     Add alternative names for isolates.	ojects to which isolates can belong. Les to projects.	isolates	+	++	query   browse	list batch update	
project members     +     +     ?     Add isolates to projects.       isolate aliases     +     ++     ?     Add isolates to projects.	tes to projects.	late field extended attribute values	; +	++		?	Add values for additional isolate field attributes.
Isolate aliases + ++ ? Add alternative names for isolates.		projects	+	++		?	Set up projects to which isolates can belong.
	native names for isolates.	project members	+	++		?	Add isolates to projects.
PubMed links + ++ 2		isolate aliases	+	++		?	Add alternative names for isolates.
T domed mins		PubMed links	+	++		?	
allele designations ++ ? Allele designations can be set within the isolate table functions.	ignations can be set within the isolate table functions.	allele designations		++		?	Allele designations can be set within the isolate table functions.
SEQUENCES ++ ? The sequence bin holds sequence contigs from any source.	ance bin holds sequence contigs from any source.	sequences		++		?	The sequence bin holds sequence contigs from any source.
accession number links + ++ ? Associate sequences with Genbank/EMBL accession number.	sequences with Genbank/EMBL accession number.	accession number links	+	++		?	Associate sequences with Genbank/EMBL accession number.
experiments + ++ ? Set up experiments to which sequences in the bin can belong.		experiments	+	++		?	Set up experiments to which sequences in the bin can belong.
experiment sequences ? Add links associating sequences to experiments.		experiment sequences				?	Add links associating sequences to experiments.
Sequence taos Scan ? Tag regions of sequences within the sequence bin with locus information.	periments to which sequences in the bin can belong.	sequence tags		scan		?	Tag regions of sequences within the sequence bin with locus information.
	periments to which sequences in the bin can belong.						
accession number links + ++ ? Associate sequences with Genbank/EMBL accession number.	sequences with Genbank/EMBL accession number.	accession number links	+	++		?	Associate sequences with Genbank/EMBL accession number.
experiments + ++ ? Set up experiments to which sequences in the bin can belong.		experiments	+	++		?	Set up experiments to which sequences in the bin can belong.
experiment sequences ? Add links associating sequences to experiments.		experiment sequences				?	Add links associating sequences to experiments.
SEQUENCE tags SCAN ? Tag regions of sequences within the sequence bin with locus information.	periments to which sequences in the bin can belong.	sequence tags		scan		?	Tag regions of sequences within the sequence bin with locus information.
	periments to which sequences in the bin can belong. associating sequences to experiments.						
	periments to which sequences in the bin can belong. associating sequences to experiments.						
	periments to which sequences in the bin can belong. associating sequences to experiments.						
	periments to which sequences in the bin can belong. associating sequences to experiments.	Database configuration	۱. I				
Catabase configuration	periments to which sequences in the bin can belong. associating sequences to experiments.		_				
Catabase configuration	periments to which sequences in the bin can belong. associating sequences to experiments. ns of sequences within the sequence bin with locus information.	Table A	dd Ba	atch Add	Update or delete		Comments
	periments to which sequences in the bin can belong. associating sequences to experiments. ns of sequences within the sequence bin with locus information.	loci -	F I	++	?		
Database configuration     Table  Add  Batch Add  Update or delete  Comments	periments to which sequences in the bin can belong. associating sequences to experiments. ns of sequences within the sequence bin with locus information.	d	ataba	ink scan			
Database configuration     Table  Add  Batch Add  Update or delete  Comments	periments to which sequences in the bin can belong. associating sequences to experiments. ns of sequences within the sequence bin with locus information.	locus aliasos	+	++	? /	Add alternative names	s for loci. These can also be set when you batch add loci.
Table     Add Batch Add Update or delete     Comments       loci     +     ++     ?	periments to which sequences in the bin can belong. associating sequences to experiments. ns of sequences within the sequence bin with locus information. Comments						· · · · · · · · · · · · · · · · · · ·
Table     Add     Batch Add     Update or delete     Comments       loci     +     ++     ?       databank scan     ?	periments to which sequences in the bin can belong. associating sequences to experiments. associating sequences to experiments. associating sequences within the sequence bin with locus information.  Comments These can also be set when you batch add loci.		1				

	Download a	submission	template i	n Excel	format	from the link.
--	------------	------------	------------	---------	--------	----------------

PTTTMLST Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Batch insert isolates	
This page allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.	
<ul> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.</li> <li>Enter references for your isolates as a semi-colon (;) separated list of PubMed ids (non integer ids will be ignored).</li> <li>You can also upload allele fields along with the other isolate data - simply create a new column with the locus name. method set as 'manual'.</li> <li>You can choose whether or not to include an id number field - if it is omitted, the next available id will be used automal</li> </ul>	
<ul> <li>Download tab-delimited header for your spreadsheet - use Paste special → text to paste the data.</li> <li>Download submission template (xlsx format)</li> </ul>	
Please select the sender from the list below:	
Select sender   Value will be overridden if you include a sender field in your pasted data.	
Paste in tab-delimited text (include a field header line).	Action
	Reset Submit

Prepare your data in the spreadsheet - the column headings must match the database fields. In databases with large numbers of loci, there won't be columns for each of these. You can, however, manually add locus columns.

Pick a sender from the drop-down list box and paste the data from your spreadsheet in to the web form. The next available isolate id number will be used automatically (this can be overridden if you manually add an id column).

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert								
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i							
Batch insert isolates								
This page allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.								
<ul> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.</li> <li>Enter references for your isolates as a semi-colon (;) separated list of PubMed ids (non integer ids will be ignored).</li> <li>You can also upload allele fields along with the other isolate data - simply create a new column with the locus name. These will be added with a confirmed status and method set as "manual".</li> <li>You can choose whether or not to include an id number field - if it is omitted, the next available id will be used automatically.</li> </ul>								
<ul> <li>Download tab-delimited header for your spreadsheet - use Paste special → text to paste the data.</li> <li>Download submission template (xlsx format)</li> </ul>								
Please select the sender from the list below:								
Jolley, Keith (keith) Value will be overridden if you include a sender field in your pasted data.								
Paste in tab-delimited text (include a field header line).								
isolate aliases references country region year epidemiological year	Reset Submit							
age_yr age_mth_sex disease source epidemiology species serogroup MLEE designation serotype sero subtype ET no penicillin								
penicillin range amoxicillin sulphonamide ceftriaxone								
ceftriaxone range chloramphenicol chloramphenicol range cefotaxime								
cefotaxime range rifampicin rifampicin range ciprofloxacin								
ciprofloxacin range pending assembly assembly status ENA accession								
private_project comments abcZ adk arcE fumC gdh pdhC								
2014 12 male								
meningitis and septicaemia CSF Neisseria meningitidis								
B								

Press submit. Data are checked for consistency and if there are no problems you can then confirm the submission.

	eith). Log out   Cha	nge passw	ord											Toggle:
tch inse <mark>r</mark> t	isolates													
oort status														
der: Keith Jolley														
obvious problems i	dentified so far.													
nport data														
ita to be import	ted													
e following table sh		nv field w	ith red text	has a problen	n and needs	to be c	hecked.							
-								0.011	diagona	0.01150.0	opidomiology	oposioo	0.050.050.00	MLEE doo
457 J392 1	es references (	UK	egion yea		gical_year	age_yr 12	age_mtn		disease meningitis	CSF	epidemiology	species Neisseria	serogroup B	MLEE_des
_									and			meningitidis		
									septicaemia					

Any problems with the data will be listed and highlighted within the table. Fix the data and resubmit if this happens.

PTIDRILST Users: Add   Query/update Isolates: Add   Query/update   Batch insert									
Logged in: Keith Jolley (keith). Log out   Change password Toggle: 1									
Batch insert isolates									
Import status									
Primary key         Problem(s)           id: 30457         species 'Neisseria meningtidis' is not on the list of allowed values for this field.									
Data to be imported									
The following table shows your data. Any field with red text has a problem and needs to be checked.									
id isolate aliases references country region year epidemiological_year age_yr age_mth sex disease	source epidemiology species serogroup MLEE_desig								
30457 J392_1 UK 2014 12 male meningitis and septicaemi	meningtidis								
<	Þ								

# 6.9 Updating and deleting single isolate records

First you need to locate the isolate record. You can either browse or use a search or list query.

tabase curator's int		ged in: Keith Jolley (Keith), @Log out   Change password								
lapase curator s int	atabase curator's interface - Neisseria PubMLST									
	eri	ace - N	eisseria Publilia							
🛃 Add, update or delete re	cord	s								
Record type A	dd B	atch Add	Update or delete	Comments						
Users	+	++	?							
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.						
	+	++	?	Add users to groups for setting access permissions.						
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.						
isolates	+	++ qu	ery   browse   list   batch update							
blate field extended attribute values	+	++	?	Add values for additional isolate field attributes.						
projects	+	++	?	Set up projects to which isolates can belong.						
project members	+	++	?	Add isolates to projects.						
isolate aliases	+	++	?	Add alternative names for isolates.						
PubMed links	+	++	?							
allele designations		++	?	Allele designations can be set within the isolate table functions.						
sequences		++	?	The sequence bin holds sequence contigs from any source.						
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.						
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.						
experiment sequences			?	Add links associating sequences to experiments.						
sequence tags	S	can	?	Tag regions of sequences within the sequence bin with locus information.						

The query interface is the same as the *public query interface*. Following a query, a results table of isolates will be displayed. There will be delete and update links for each record.

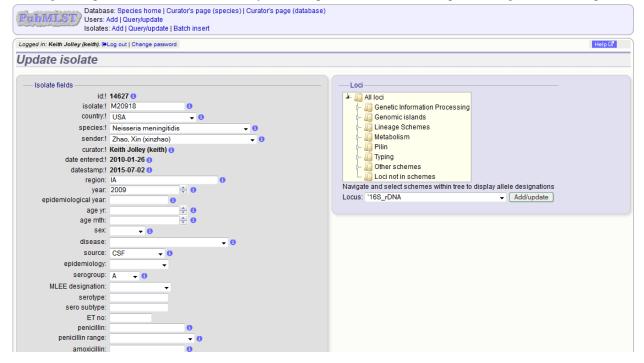
Leiete       update       upload       create       12674       M18700       USA       2009       Neisseria meningitidis       B       0       7301       ST-32 complex/ET-5 complex/ET-5       add         Leiete       update       upload       create       12675       M18701       USA       2009       Neisseria meningitidis       B       0       7302       ST-32 complex/ET-5 complex       add         Leiete       update       upload       create       12676       M18705       USA       2009       Neisseria meningitidis       B       0       7303       add       add         Leiete       update       upload       create       13090       M19024       PA09015       USA       2009       meningitidis       Neisseria meningitidis       B       0       7575       Leiete       add         Leiete       update       update       update       Leiete       USA       2009       meningitidis       Neisseria meningitidis       A       1718070       7890       ST-52       20         Leiete       update       update       Leiete       USA       2009       Meningitidis       A       1718070       7890       ST-52       20       20       20       Meningitidi	gged ir	: Keith Jo	lley (keith). L	.og out   Cł	nange pa	ssword					То	ggle: i Fie	ld help: id			-	G	
Combine with:       AND       Action       Order by:       id       Image: Complex with and	ola	te qu	ery/up	odate														
Combine with:         AND         Image: Combine with:         Order by:         id         Image: Combine with:         Second and and and and and and and and and a	Iso	late prove	enance/phe	notype fie	elds —					Display/sort optic	ons —						lod	
year       Image: second	Com	oine with:	AND 👻							Order by: id			•	ascend	ling 👻			
Action       Section	cou	ntry	-	=	•	- USA		+ [	i	Display: 25		r page 👔				0		
Reset Submit         Submit         records returned. Click the hyperlinks for detailed information.         Delete       Tag scanning       Projects         Delete ALL       Scan       Select project       Link         New bin       MLST       find isolate       aliases       country year       disease       species       serogroup       Size (ho)         MLST       find isolate       aliases       country year       disease       species       serogroup       Size (ho)         MLST       find isolate       aliases       country year       disease       species       serogroup       Size (ho)         MLST       disease       species       serogroup       Size (ho)         Idelete update       upload       create       12676       MI870       USA       200       Neisseria       B       0 <th colsp<="" td=""><td>yea</td><td>r</td><td>-</td><td>=</td><td>•</td><td>2009</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th>	<td>yea</td> <td>r</td> <td>-</td> <td>=</td> <td>•</td> <td>2009</td> <td></td>	yea	r	-	=	•	2009											
Reset       Submit         records returned. Click the hyperlinks for detailed information.         Delete         Tag scanning         Projects         Delete ALL         Sequenc       New         Select project         Link         Setup date       MLST       Find         Joint colspan="2">Setup date       New         Setup date       New       Select project       Link         Setup date       New       Select project       Link         Interview date       New       Select project       Link         Setup date       New       Select project       Link         Interview date       New       Select project       Link         Projects       Select project       Link         Setup date       New       Setup date       Neisseria       B       0       Tot conal complex       PorA         Projects       Setup date       Upload       create       12674       M18700       USA       2009       Neisseria       B       0	Act	ion																
records returned. Click the hyperlinks for detailed information. Delete Tag scanning Projects Delete ALL Scan Select project Link elete Update Update upload create 12674 M18700 USA 2009 Neisseria B 0 7301 Conal complex Port of the select project and the select project of the s	10																	
Delete       Tag scanning       Projects         Delete ALL       Scan       Select project       Link	Re	set S	Submit															
Delete       Tag scanning       Projects         Delete ALL       Scan       Select project       Link																		
Delete       Tag scanning       Projects         Delete ALL       Scan       Select project       Link         elete       Update       Sequence       New       Select project       Link         elete       update       upload       create       12674       M18700       USA       2009       Neisseria       B       0       7301       ST-32       add         elete       update       upload       create       12675       M18701       USA       2009       Neisseria       B       0       7301       ST-32       add         elete       update       upload       create       12676       M18701       USA       2009       Neisseria       B       0       7302       complex/ET-5       complex/ET-5       complex/ET-5       complex/ET-5       complex/ET-5       add       add <t< th=""><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></t<>																		
Delete       Tag scanning       Projects         Delete ALL       Scan       Select project       Link         elete       Update       Sequence       New       Isolate       aliases       country year       disease       species       serogroup       Setophone       MLST       Fin are         elete       update       upload       create       12674       M18700       USA       2009       Neisseria meningitidis       B       0       7301       ST-32 complex/ET-5 complex/ET-5 complex/ET-5 complex       add         elete       update       upload       create       12676       M18701       USA       2009       Neisseria meningitidis       B       0       7301       ST-32 complex/ET-5 complex       add         elete       update       upload       create       12676       M18705       USA       2009       Neisseria meningitidis       B       0       7303       ST-32 complex/ET-5 complex       add         elete       update       upload       create       12676       M18725       USA       2009       Neisseria meningitidis       B       0       7575       add         elete       update       upload       create       13090       M19024       PA09015<	recor	de roturn	ad Click th	e hyperlin	ke for d	etailed inform	nation											
Delete ALL       Scan       Select project       Link         New bin       New version	recor	us return	ieu. Ciick ui	e nypenn	iks lui u	etalleu mon	nauon.											
Delete ALL       Scan       Select project       Link         elete       Update       Sequence       New version       id       isolate       aliases       country year       disease       species       serogroup       size (bp)       ST       clonal complex       PorA         elete       update       upload       create       12674       M18700       USA       2009       Neisseria       B       0       7301       ST-32       add         elete       update       upload       create       12675       M18701       USA       2009       Neisseria       B       0       7302       complex/ET-5       complex/ET-5       complex/ET-5       complex/ET-5       complex/ET-5       add         elete       update       upload       create       12676       M18725       USA       2009       Neisseria       B       0       7303       add	Del																	
elete       Update       Sequence bin       New version       id       isolate       aliases       country       year       disease       species       serogroup       Size (bp)       ST       clonal complex       PorA VR1         elete       update       upload       create       12674       M18700       USA       2009       Neisseria meningitidis       B       0       7301       ST-32 complex/ET-5 complex       add         elete       update       upload       create       12675       M18701       USA       2009       Neisseria meningitidis       B       0       7302       ST-32 complex/ET-5 complex/ET-5 complex       add         elete       update       upload       create       12676       M18705       USA       2009       Neisseria meningitidis       B       0       7303       add       add         elete       update       upload       create       12676       M18725       USA       2009       Neisseria meningitidis       B       0       7575       add       add         elete       update       upload       create       1309       M19024       PA09015       USA       2009       Neisseria meningitidis       B       0       7575		ete ———	— — Tao s	canning	PI	roiects												
Verter       New bin       New version       New version <th< td=""><td></td><td></td><td></td><td></td><td>_</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th<>					_													
Perfecte       Update       Sequence bin       New version       Neisseria meningitidis       B       0       7302       ST-32 complex/ET-5 complex/E	Dele				_				▼ L	ink								
Verter       New bin       New version       New version <th< th=""><th>Dele</th><th></th><th></th><th></th><th>_</th><th></th><th></th><th></th><th>▼ L</th><th>ink</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></th<>	Dele				_				▼ L	ink								
Sequence bin       New version       id       isolate       aliases       country year       disease       species       serogroup       size (b)       ST       clonal complex       Vrt1       Vrt1         lelete       update       upload       create       12674       M18700       USA       2009       Neisseria meningitidis       B       0       7301       ST-32 complex/ET-5 complex/ET-5 complex/ET-5 complex/ET-5       add         lelete       update       upload       create       12676       M18701       USA       2009       Neisseria meningitidis       B       0       7302       ST-32 complex/ET-5 complex/ET-5       add         lelete       update       upload       create       12676       M18705       USA       2009       Neisseria meningitidis       B       0       7303       add       add         lelete       update       upload       create       12676       M18725       USA       2009       meningitidis       B       0       7303       add       add         lelete       update       upload       create       13090       M19024       PA09015       USA       2009       meningitidis       A       1718070       4789,       ST-5       20 <th>Dele</th> <th></th> <th></th> <th></th> <th>_</th> <th></th> <th></th> <th></th> <th>▼ L</th> <th>ink</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>	Dele				_				▼ L	ink								
Initial contract       Version       id       isolate       aliases       country year       disease       species       serogroup       SIZe (6)       ST       clonal complex       PorA         lelete       update       upload       create       12675       M18700       USA       2009       Neisseria meningitidis       B       0       7301       ST-32 complex/ET-5 complex/ET-5 complex       add         lelete       update       upload       create       12675       M18701       USA       2009       Neisseria meningitidis       B       0       7302       ST-32 complex/ET-5 complex       add         lelete       update       upload       create       12676       M18701       USA       2009       Neisseria meningitidis       B       0       7302       add         lelete       update       upload       create       12676       M18725       USA       2009       Meningitidis       B       0       7303       add         lelete       update       upload       create       13090       M19024       PA09015       USA       2009       meningitidis       B       0       7575       add         lelete       update       update       update <t< th=""><th>Dele</th><th></th><th>S</th><th>can</th><th>_</th><th></th><th></th><th>150</th><th></th><th></th><th></th><th></th><th></th><th></th><th>MI ST</th><th></th><th></th></t<>	Dele		S	can	_			150							MI ST			
lefete       update       upload       create       12675       M18701       USA       2009       Neisseria meningitidis       RB       00       7302       complex/ET-5 complex       add         lefete       update       upload       create       12675       M18701       USA       2009       Neisseria meningitidis       RB       00       7303       complex/ET-5 complex       add         lefete       update       upload       create       12676       M18725       USA       2009       Neisseria meningitidis       RB       00       7303       complex/ET-5 complex       add         lefete       update       upload       create       1309       M19024       PA09015       USA       2009       meningitidis       Neisseria meningitidis       RB       0       7575       add         lefete       update       upload       create       14627       M20918       USA       2009       Meningitidis       A       1718070       4789.       Complex/Subgroup       update       110		ete ALL	Sequence	New	_			Iso							MLST	á	an	
Lefete       updad       create       12675       M18701       USA       2009       Neisseria meningitidis       B       0       7302       complex       add         Lefete       updad       upload       create       12675       M18725       USA       2009       Neisseria meningitidis       B       0       7302       add       add         Lefete       updad       upload       create       12676       M18725       USA       2009       meningitidis       neiningitidis       0       7303       add       add         Lefete       updad       upload       create       13090       M19024       PA09015       USA       2009       meningitidis       Neisseria meningitidis       0       7575       add       add         Lefete       updad       upload       create       14627       M20918       USA       2009       meningitidis       A       1718070       4789.       Complex/subgroup       update       update       update       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1118070       1		ete ALL	Sequence	New	Sel	ect project	aliases	1	ate field	s i	species	serogroup		ST		ة PorA	an	
Lelete       upload       create       12675       M18701       USA       2009       Neisseria meningitidis       B       0       7302       add         Lelete       upload       create       12676       M18725       USA       2009       Neisseria meningitidis       B       0       7302       add       add         Lelete       upload       create       13090       M18725       USA       2009       meningitidis       B       0       7303       add         Lelete       upload       create       13090       M19024       PA09015       USA       2009       meningitidis       Neisseria meningitidis       B       0       7575       add       add         Lelete       upload       create       14627       M20918       USA       2009       Meningitidis       A       1718070       4789,       5T-5       20         update       upload       create       14627       M20918       USA       2009       Neisseria meningitidis       A       1718070       4789,       5T-5       20         update       update       update       update       update       10       10       10       10       10       10       10	elete	update	Sequence bin	New version	Sel	ect project isolate	aliases	country	ate field year	s i			size (bp)		clonal complex	PorA VR1	an	
Index       Index <th< td=""><td>elete</td><td>update</td><td>Sequence bin</td><td>New version</td><td>Sel</td><td>ect project isolate</td><td>aliases</td><td>country</td><td>ate field year</td><td>s i</td><td>Neisseria</td><td></td><td>size (bp)</td><td></td><td>clonal complex ST-32</td><td>PorA VR1</td><td>an</td></th<>	elete	update	Sequence bin	New version	Sel	ect project isolate	aliases	country	ate field year	s i	Neisseria		size (bp)		clonal complex ST-32	PorA VR1	an	
lelete       upload       create       12676       M18725       USA       2009       Neisseria       B       0       7303       add         lelete       upload       create       13090       M19024       PA09015       USA       2009       meningitidis       Neisseria       B       0       7575       add         lelete       upload       create       14627       M20918       USA       2009       meningitidis       Neisseria       B       0       7575       add         upload       create       14627       M20918       USA       2009       Neisseria       A       1718070       4789       ST-5       20         upload       create       14627       M20918       USA       2009       Neisseria       A       1718070       4789       ST-5       20         upload       initia       initia       initia       initia       initia       initia       initia       initia	elete	update	Sequence bin	New version	Sel	ect project isolate	aliases	country	ate field year	s i	Neisseria		size (bp)		cional complex ST-32 complex/ET-5	PorA VR1	an	
Lelete     upload     create     13090     M19024     PA09015     USA     2009     meningitidis     Neisseria meningitidis     B     0     7575     add       Lelete     upload     create     14627     M20918     USA     2009     meningitidis     Neisseria meningitidis     A     1718070     4789, 7980     ST-5     20       upload     create     14627     M20918     USA     2009     Neisseria meningitidis     A     1718070     4789, 7980     ST-5     20       upload     III     III     III     III     III     III     III     III	elete lelete	Update	Sequence bin upload	New version create	Sel id 12674	ect project isolate M18700	aliases	country USA	ate field year 2009	s i	Neisseria meningitidis	В	size (bp) O	7301	cional complex ST-32 complex/ET-5	PorA VR1 add	an	
lelete       upload       create       13090       M19024       PA09015       USA       2009       meningitits       Neisseria       B       0       7575       add         lelete       upload       create       14627       M20918       USA       2009       Meningitits       Neisseria       A       1718070       4789.       ST-5       20         upload       create       14627       M20918       USA       2009       Neisseria       A       1718070       4789.       ST-5       20         upload       upload       create       14627       M20918       USA       2009       Neisseria       A       1718070       4789.       Complex/subgroup       update       10	elete lelete	Update	Sequence bin upload	New version create	Sel id 12674	ect project isolate M18700	aliases	country USA	ate field year 2009	s i	Neisseria meningitidis Neisseria	В	size (bp) O	7301	cional complex ST-32 complex/ET-5	PorA VR1 add	an I	
lelete update upload create 14627 M20918 USA 2009 Neisseria A 1718070 4789, ST-5 20 meningitidis Marcine M Marcine Marcine Mar	elete lelete lelete	Update update update	Sequence bin upload upload	New version create create	Sel id 12674 12675	isolate M18700	aliases	Country USA USA	ate field year 2009 2009	s i	Neisseria meningitidis Neisseria meningitidis	В	size (bp) O	7301 7302	cional complex ST-32 complex/ET-5	e PorA VR1 add add	an I	
elete update upload create 14627 M20918 USA 2009 Neisseria A 1718070 4789, ST-5 20 meningitidis 1718070 4789, ST-5 20 111	elete elete elete	Update update update	Sequence bin upload upload	New version create create	Sel id 12674 12675	isolate M18700	aliases	Country USA USA	ate field year 2009 2009	s i	Neisseria meningitidis Neisseria meningitidis Neisseria	В	size (bp) O	7301 7302	cional complex ST-32 complex/ET-5	e PorA VR1 add add	an I	
meningitidis 7980 complex/subgroup update i	elete elete elete	Update Update Update Update	Sequence bin upload upload upload	New version create create create	id 12674 12675 12676	isolate M18700 M18701 M18725		Country USA USA USA	ate field year 2009 2009 2009	s 🛿	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B	size (bp) 0 0 0	7301 7302 7303	cional complex ST-32 complex/ET-5	e PorA VR1 add add add	an	
meningitidis 7980 complex/subgroup update i	elete elete elete	Update Update Update Update	Sequence bin upload upload upload	New version create create create	id 12674 12675 12676	isolate M18700 M18701 M18725		Country USA USA USA	ate field year 2009 2009 2009	s 🛿	Neisseria meningitidis Neisseria meningitidis Neisseria Neisseria	B	size (bp) 0 0 0	7301 7302 7303	cional complex ST-32 complex/ET-5	e PorA VR1 add add add	an	
	)elete delete delete delete delete	Update Update Update Update Update	Sequence bin upload upload upload upload	New version create create create create	Sel id 12674 12675 12676 13090	ect project isolate M18700 M18701 M18725 M19024		Country USA USA USA USA	ate field year 2009 2009 2009 2009	s 🛿	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	BBB	size (bp) 0 0 0 0	7301 7302 7303 7575	clonal complex ST-32 complex/ET-5 complex	PorA VR1 add add add add	an  I	
	)elete delete delete delete delete	Update Update Update Update Update	Sequence bin upload upload upload upload	New version create create create create	Sel id 12674 12675 12676 13090	ect project isolate M18700 M18701 M18725 M19024		Country USA USA USA USA	ate field year 2009 2009 2009 2009	s 🛿	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria	BBB	size (bp) 0 0 0 0 0 0	7301 7302 7303 7575 4789,	Clonal complex ST-32 complex/ET-5 complex	PorA VR1 add add add add 20		
lelete update upload create 14998 M21319 USA 2009 invasive Neisseria B 0 8250 ST-60 complex add	elete lelete lelete lelete lelete	Update Update Update Update Update	Sequence bin upload upload upload upload	New version create create create create	Sel id 12674 12675 12676 13090	ect project isolate M18700 M18701 M18725 M19024		Country USA USA USA USA	ate field year 2009 2009 2009 2009	s 🛿	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria	BBB	size (bp) 0 0 0 0 0 0	7301 7302 7303 7575 4789,	Clonal complex ST-32 complex/ET-5 complex	PorA VR1 add add add add 20		

Clicking the 'Delete' link takes you to a page displaying the full isolate record.

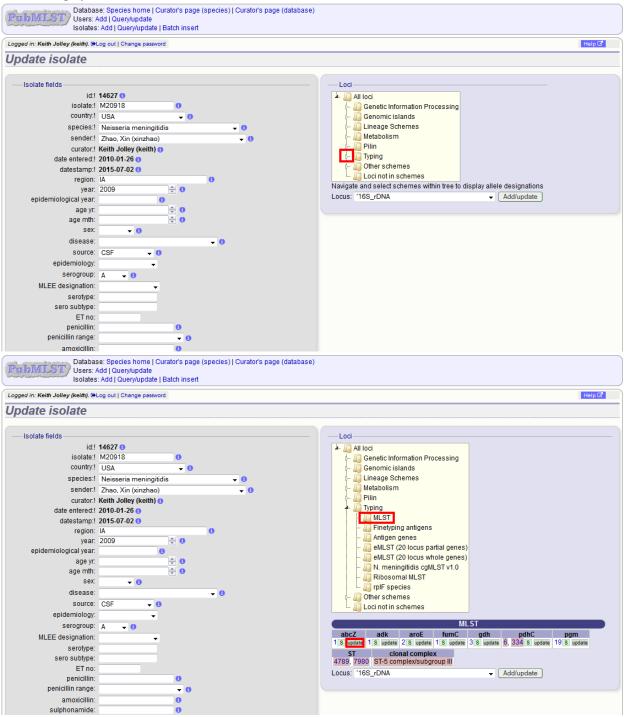
PTITIVIUST Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert							
Logged in: Keith Jolley (ke	eith). Log out   Change password						
Delete isolat	6						
You have selected to d	lelete the following record:						
Provenance/meta	data						
id:	14627	source:	CSF	update history:	26 updates show details		
isolate:	M20918	species:	Neisseria meningitidis	date entered:	2010-01-26		
strain designation:	A: P1.20,9: F3-1: ST-4789,7980	serogroup:	Α	datestamp:	2014-06-17		
	(cc5)	sender:	Xin Zhao, Novartis (formerly at US				
country:			CDC)				
	North America	curator:	Carina Brehony, University of Oxford,				
region:			UK (E-mail: carina.brehony@zoo.ox.ac.uk)				
year:	2009		canna.breneny@200.0x.ac.uky				
Publication (1)							
			diti V, Humphrey JC, Sammons SA, Govil acing projects. <i>Bioinformatics</i> 26:1819-26		, Tondella ML, Harcourt BH, Mayer LW,		
Sequence bin							
configs:	2049	N90:	456				
-		N95:					
-		loci tagged:	979				
-		letailed breakdown:					
	1155						
A - N							
Action							
Delete							
Jordan IK (201 Sequence bin total length: max length: mean length: N50: Action	0). A computational genomics pipeline f 2049 1718070 bp 5436 bp 839 bp	for prokaryotic sequer N90: N95: loci tagged:	acing projects. <i>Bioinformatics</i> 26:1819-26 456 324 979		, concerne me, riarcourteri, mayor Err,		

Pressing 'Delete' from this record page confirms the deletion.

Clicking the 'Update' link for an isolate takes you to an update form. Make the required changes and click 'Update'.



Allele designations can also be updated by clicking within the scheme tree and selecting the 'Add' or 'Update' link



#### next to a displayed locus.

Schemes will only appear in the tree if data for at least one of the loci within the scheme has been added. You can additionally add or update allelic designations for a locus by choosing a locus in the drop-down list box and clicking 'Add/update'.

Patabase: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert							
Logged in: Keith Jolley (keith).	Log out   Change password	Help 🗹					
Update isolate							
Isolate fields	4.62	Lod					
	14627 0 M20918 0	A 🗐 All loci					
country:		- 💭 Genetic Information Processing - 💭 Genomic islands					
-	Neisseria meningitidis	- Lineage Schemes					
	Zhao, Xin (xinzhao)	- Metabolism					
	Keith Jolley (keith) ()	- 🛺 Pilin					
	2010-01-26 0	b 🛺 Typing					
datestamp:	2015-07-02 ()	- 🛺 Other schemes					
region:	IA 🕚	Loci not in schemes					
year	2009 🔷 🕄 🚺	Navigate and select schemes within tree to display allele designations					
epidemiological year:		Locus: abcZ    Add/update					
age yr.							
age mth:							
sex	<b>•</b> 0						
disease	• 0						
source	CSF 🗸 🚺						
epidemiology	<b>•</b>						
serogroup	A 👻 🖲						
MLEE designation:	<b>•</b>						
serotype							
sero subtype:							
ET no:							
penicillin							
penicillin range:							
amoxicillin:							
sulphonamide:	0						

The allele designation update page allows you to modify an existing designation, or alternatively add additional designations. The sender, status (confirmed/provisional) and method (manual/automatic) needs to be set for each designation (all pending designations have a provisional status). The method is used to differentiate designations that have been determined manually from those determined by an automated algorithm.

(PubMLST) U	Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert						
Logged in: Keith Jolley (ke	eith). Log out   Change password						
Update abcZ	allele for isolate 1462	27					
Provenance/meta data		Locus: abcZ					
id:	14627	Add new allele designation					
isolate:	M20918	Please fill in the fields below - required fields are marked with an exclamation mark (!).					
strain designation:	A: P1.20,9: F3-1: ST-4789,7980						
	(cc5)	Record					
country:		isolate id:! 14627					
region:	2009	locus:! abcZ allele id: 15					
source:							
	Neisseria meningitidis	status: I confirmed -					
serogroup:	-	method: manual					
	Xin Zhao	curator:! Keith Jolley (keith)					
	Carina Brehony	datestamp:/ 2014-07-10					
	26 updates Show details	date entered:! 2014-07-10					
date entered:		comments:					
datestamp:	2014-06-17	Action					
Update other loci	:	Reset Submit					
Locus: abcZ	✓ Add/update	Existing designations					
		Update         Delete         allele id         sender         status         method         comments           Update         Delete         1         Xin Zhao         confirmed         manual					

## 6.10 Batch updating multiple isolate records

Database: Species home | Curator's page (species) | Curator's page (database) PubMLS' Users: Add | Query/update Isolates: Add | Query/update | Batch insert Logged in: Keith Jolley (keith). Hog out | Change password Database curator's interface - Neisseria PubMLST 🔗 Add, update or delete records Record type Add Batch Add Update or delete Co users Users can be members of these groups - use for setting access permit user groups + ++ user group members Add users to groups for setting access permissions curator permissions Set o missions for individual users - these ar query | browse | list batch update isolates isolate field extended attribute values + Add values for additional isolate field attributes projects ++ Set up projects to which isolates can belong. ++ project members Add isolates to projects. isolate aliases Add alternative names for isolates PubMed links ++ allele designations Allele designations can be set within the isolate table functions ++ sequences The sequence bin holds sequence contigs from any source. accession number links ++ Associate sequences with Genbank/EMBL accession number ++ Set up experiments to which sequences in the bin can belong. experiments experiment sequences Add links associating sequences to experiment scan sequence taos Tag regions of sequences within the sequence bin with locus information Database configuration Add Batch Add Update or delete Table loci ? databank scan locus aliases Add alternative names for loci. These can also be set when you batch add loci PCR reactions ++ Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci nucleotide probes ++ Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetit isolate field extended attributes ++ Define additional attributes to associate with values of a particular isolate record field

Select 'batch update' isolates link on the curator's index page.

Prepare your update data in 3 columns in a spreadsheet:

- 1. Unique identifier field
- 2. Field to be updated
- 3. New value for field

You should also include a header line at the top - this isn't used so can contain anything but it should be present.

Columns must be tab-delimited which they will be if you copy and paste directly from the spreadsheet.

So, to update isolate id-100 and id-101 to serogroup B you would prepare the following:

```
id field value
100 serogroup B
101 serogroup B
```

Select the field you are using as a unique identifier, in this case id, from the drop-down list box, and paste in the data. If the fields already have values set, you should also check the 'Update existing values' checkbox. Press 'submit'.

Putitivities: Add   Query/update Isolates: Add   Query/update   Batch insert							
Logged in: Keith Jolley (keith). Log out   Change password							
Batch isolate update							
This page allows you to batch update provenance fields or allele	designations for multiple isolates.						
<ul> <li>The first line, containing column headings, will be ignored.</li> <li>The first column should be the isolate id (or unique field that you are selecting isolates on). If a secondary selection field is used (so that together the combination of primary and secondary fields are unique), this should be entered in the second column.</li> <li>The next column should contain the field/locus name and then the final column should contain the value to be entered, e.g.</li> </ul>							
id field value							
2 country USA 2 abcZ 5							
<ul> <li>The columns should be separated by tabs. Any other colu         <ul> <li>If you wish to blank a field, enter '<blank>' as the value.</blank></li> <li>The script is compatible with STARS output files.</li> </ul> </li> <li>Please enter the field(s) that you are selecting isolates on. Value the database id will be used.</li> </ul>		is, i.e. only one isolate has the value(s) used. Usually					
Please paste in your data below:	- Options	Allele designations					
id field value	Primary selection field: id	Add additional new designation					
100 serogroup B 101 serogroup B	Optional selection field: <none></none>	Replace existing designations					
101 serogroup B	Update existing values	Action					
		Reset Submit					
j.							

A confirmation page will be displayed if there are no problems. If there are problems, these will be listed. Press 'Upload' to upload the changes.

PubML	Prink/ILST Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert								
Logged in: Keith	Logged in: Keith Jolley (keith). Log out   Change password								
Batch is	Batch isolate update								
The following	changes will be m	ade to th	e database. Please check that th	this is what you intend and then press 'Upload'. If you do not wish to make these changes, press your					
browser's bac	k button.								
Transaction	id Field N	ew value	Value(s) currently in database	e Action					
	100 serogroup	В		update field with new value					
	101 serogroup	В	С	update field with new value					
Action Upload Back to main	page								

You can also use a secondary selection field such that a combination of two fields uniquely defines the isolate, for example using country and isolate name.

So, for example, to update the serogroups of isolates CN100 and CN103, both from the UK, select the appropriate primary and secondary fields and prepare the data as follows:

isolate	country	field	value
CN100	UK	serogroup	В
CN103	UK	serogroup	В

# 6.11 Deleting multiple isolate records

**Note:** Please note that standard curator accounts may not have permission to delete multiple isolates. Administrator accounts are always able to do this.

Before you can delete multiple records, you need to search for them. From the curator's main page, click the Query isolates link:

Database: Spec Users: Add   Qu Isolates: Add   Q	ery/up	odate	tor's page (species)   Curato tch insert	rs page (database)
gged in: Keith Jolley (keith). ⊯Log out	Chang	e password		
atabase curator's i	nto	rfaco	Noissoria Pub	T2 IM
alabase curators i	ne	ilace -	NCISSCIIA FUN	
Add. update or delete	rees	arde		
Record type	Add	Batch Add		Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'ourator' in the users table.
isolates	+	++	query browse   list   batch	update
olate field extended attribute value	s +	++	?	Add values for additional isolate field attributes.
projects	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
isolate aliases	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags		scan	?	Tag regions of sequences within the sequence bin with locus information.
Database configuration				
Table A	dd Ba	atch Add U	pdate or delete	Comments
1001	+ lataba	++ ank scan	?	
	+	++	? Add alternativ	ve names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		o PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
	+	++		otide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
			Denne hudet	and proves for in one or interaction resident to their genomes for lagging to specific repetitive rout.

Enter search criteria that specifically return the isolates you wish to delete. Click 'Delete ALL'.

nor show	-t	Isolates	. Add I Qu	eryapus	and Logici	moon												
Logged in: Keith Jolley (keith). @Log out   Change password										H	elp 🗗 🛛 Te	oggle: 🚯 🛛 F	ield help:	id				-
sola	te qu	ery/up	date															
							— Display/sort op								Modi			
Comb	ine with:	AND 👻							Order by: id	Order by: id				→ ascending →         →         →         →				
date	entered	-	-	•	2014-03	-17		+ 🚺	Display: 25	5 👻 records per	page 🕦							optio
cura	tor (surn	ame) 👻	=	-	Jolley													_
— Acti	on ——																	
Res	at C	Bubmit																
Res	et c																	
		Jubilit																
record	s returne	d. Click the I	typerlinks	for deta	iled inforr	mation.												
record:		d. Click the I				mation.												
- Dele	te	d. Click the I — — Tag si	canning-	— Pro	jects —				Link									
- Dele		d. Click the I	canning-	— Pro				•	Link									
- Dele	te	d. Click the I — — Tag si	canning-	— Pro	jects —			•	Link									
- Dele	te	d. Click the I — — Tag si Sc	canning— an	— Pro	jects —				Link e fields ()			Contrin			MLST	Finety	ping and	igens
Dele	te	d. Click the I — — Tag si	canning-	— Pro	jects —		country			species	serogroup	Seqbin size (bp)	Contigs	ST	MLST clonal complex	Finety PorA VR1	ping ant PorA VR2	igens FetA VR
Dele	te —	d. Click the I — Tag si Sc Sequence	canning— an New	- Pro Selec	ojects —— ct project.		Czech	Isolat	e fields ()	Neisseria	serogroup W		Contigs	ST 10733		PorA	PorA	FetA
Dele Delete delete	te ALL Update update	d. Click the I — Tag su Sc Sequence bin upload	canning – an New version create	Pro Select id 28783	ojects ct project. isolate 0012/14		Czech Republic	Isolat year 2014	e fields <b>()</b> disease carrier	Neisseria meningitidis	W	size (bp) O	0	10733	cional complex	PorA VR1 5-2 update	PorA VR2 10-1 update	FetA VR F5-8 update
Dele Delete delete	te ALL Update	d. Click the I — — Tag sı Sc Sequence bin	canning – an New version	Pro Select id 28783	ojects		Czech Republic Czech	Isolat year	e fields <b>()</b> disease	Neisseria meningitidis Neisseria		size (bp)	_			PorA VR1 5-2 update 22-1	PorA VR2 10-1 update 14	FetA VR F5-8 update F4-1
Delete Delete delete delete	te ALL Update update	d. Click the I — Tag su Sc Sequence bin upload	canning – an New version create	Pro Selea id 28783 28784	ojects ct project. isolate 0012/14		Czech Republic Czech Republic	Isolat year 2014	e fields <b>()</b> disease carrier	Neisseria meningitidis	W	size (bp) O	0	10733	cional complex	PorA VR1 5-2 update	PorA VR2 10-1 update	FetA VR F5-8 update

You will have a final chance to change your mind:

PubML	Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert							
Logged in: Keith Jolley (keith). II+Log out   Change password								
Delete multiple isolate records								
Warning-								
	If you proceed, you will delete 3 isolate records. Please confirm that this is your intention.							

Click 'Confirm deletion!'.

# 6.12 Uploading sequence contigs linked to isolate records

### 6.12.1 Select isolate from drop-down list

To upload sequence data, click the sequences batch add (++) link on the curator's main page.

ged in: Keith Jolley (keith). 🗭 Log out					
tabase curator's i	inte	rface -	Neisser	ia PubN	ILST
🝠 Add, update or delete	e reco	ords			
-		Batch			
Record type	Add	Add	Update	or delete	Comments
users	+	++		?	
user groups	+	++		?	Users can be members of these groups - use for setting access permissions.
user group members	+	++		?	Add users to groups for setting access permissions.
curator permissions				?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the user
					table.
isolates	+	++		owse   list   update	
isolate field extended attribute values	+	++		?	Add values for additional isolate field attributes.
projects	+	++		?	Set up projects to which isolates can belong.
project members	+	++		?	Add isolates to projects.
isolate aliases	+	++		?	Add alternative names for isolates.
PubMed links	+	++		?	
allele designations		++		?	Allele designations can be set within the isolate table functions.
sequences		++		?	The sequence bin holds sequence contigs from any source.
accession number links	+	++		?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++		?	Set up experiments to which sequences in the bin can belong.
experiment sequences				?	Add links associating sequences to experiments.
sequence tags		scan		?	Tag regions of sequences within the sequence bin with locus information.
💆 Database configuratio	on				
Table	Add Ba	atch Add U	pdate or delete		Comments
loci	+	++	?		
	databa	ank scan			
locus aliases	+	++	?	Add alternative	names for loci. These can also be set when you batch add loci.

Select the isolate that you wish to link the sequence to from the dropdown list box. You also need to enter the person who sent the data. Optionally, you can add the sequencing method used.

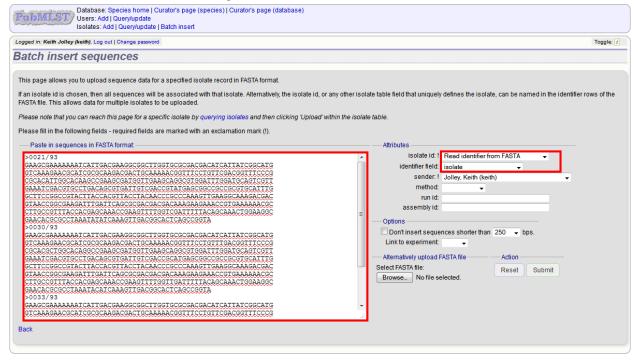
#### Paste sequence contigs in FASTA format in to the form.

Profit MJLST Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Batch insert sequences	
This page allows you to upload sequence data for a specified isolate record in FASTA format. If an isolate id is chosen, then all sequences will be associated with that isolate. Alternatively, the isolate id, or any other isol FASTA file. This allows data for multiple isolates to be uploaded.	ate table field that uniquely defines the isolate, can be named in the identifier rows of the
Please note that you can reach this page for a specific isolate by querying isolates and then clicking 'Upload' within the isolat Please fill in the following fields - required fields are marked with an exclamation mark (!).	le table.
Paste in sequences in FASTA format TITRATICTAAAATATATTITTITTCAAAGTTACTCATATTTAGTTAGTTTCCTTTTACT GGGTAIGTAAACAAATTACCTGGTAATACTGTAATTAGTTGTGGTGGGT TGTCCACCTTCATTAATAGAAGTAGTAGTACCAATAACTTTTCCTACACACGACAGTCCGCATA TATTGGCCATCAGGAGGAGTGGTAAAAACTGAACGGACAGTATTCGCAATAGGCCATAGG AAGTGCCCCTCAAGACAGCAGCAGCAGGAGAAGAACTGACCGGTTGTCCCCATAGGCCATAG GGTGTATATCGAGTAGTACTAATCGCTGTAATTAGTGTGTTTACAGCACCATAATTT	Attributes
TIGATICITETAGTAGCATCCAGTICTATACTETCTGGGTTGTATAGCCTGCTTG SCCGCCTTCGCTTCGCACAGCCCCCCACGGCGCTCCGCCCACGGGTGGTGCTACC GCACCCCGGGCGATTCCCCCCACGCGCCGCTTCGCCCACGCTTACC GCACCTCCAGCATAGCCGTTAGCGGTAATTACCCCCTGTGATGCCATA CGGATATTGCTTATCCAATGGCAGCATCCTTATCCTGCGGATTAGTCATCGCACCTGC GCACGTGCAGGCATAATCTTGGTGATATCACAATCGGGCGGTGGTGCCATA CGGATATTGCTTATCCAATGGCAGCATCCTTATCCTGCGGATTAGTCATCGCACCTGC GCAGTGCAGGCATAATCTTGGTGATTTTCCCCGGCTGGTGGCGCCGTCG GCAGTGCAGGCATAATCTGGGTAACTTTACCCAGTGCGGTTAGTCCACCCGC GCAGTGCAGGCATAATCTGGCGCAGTATTCCACCAGTGCGGTTAGTCCACCGCC GCCACCTCAGCCGCAGTATGGTAGGGTACCGTCACATCCACCGCCGCCGCGCGCG	Options Don't insert sequences shorter than 250 • bps. Link to experiment Alternatively upload FASTA file Action Select FASTA file: Browse_ No file selected.
Back	

ged in: Keith Jolley (keith). Log out   Change password			Toggle:
atch insert sequences			
ne following sequences will be entered.			
Original designation S	equence length Comm	ients	
180426 NODE_1211_length_26135_cov_16.536217	26167	<ul> <li>Number of contigs: 359</li> </ul>	
180427 NODE_619_length_1208_cov_9.455298	1240	Minimum length: 265     Maximum length: 23552	
180428 NODE_655_length_1534_cov_15.418513	1566	Maximum length: 33563     Total length: 2059411	
180429 NODE_675_length_1548_cov_17.753876	1580	Mean length: 5736	
180430 NODE_31_length_8390_cov_14.525030	8422	N50: 10581	
180431 NODE_254_length_2721_cov_18.400587	2753	• N90: 3126	
180432 NODE_262_length_1955_cov_8.388747	1987	• N95: 1863	
80433 NODE_252_length_30791_cov_15.767627	30823		
180434 NODE_189_length_536_cov_31.078358	568	Upload	
180435 NODE_716_length_511_cov_12.113503	543		
80436 NODE_465_length_13739_cov_15.131669	13771		
180437 NODE_38_length_15888_cov_17.174660	15920		
180438 NODE_778_length_2398_cov_8.673060	2430		
180439 NODE_1765_length_657_cov_8.754947	689		
180440 NODE_729_length_255_cov_13.007843	287		
180441 NODE_52_length_16337_cov_17.192997	16369		
180442 NODE_190_length_652_cov_116.434052	684		
180443 NODE_95_length_3094_cov_14.927279	3126		
180444 NODE_770_length_5072_cov_12.878943	5104		
180445 NODE_263_length_1223_cov_10.451349	1255		
180446 NODE_181_length_4496_cov_14.635231	4528		
180447 NODE_558_length_4436_cov_14.670198	4468		
180448 NODE_1179_length_917_cov_13.504908	949		
180449 NODE_527_length_4033_cov_15.608480	4065		
180450 NODE_212_length_2435_cov_12.223409	2467		
180451 NODE_54_length_790_cov_9.398734	822		
80452 NODE_146_length_12761_cov_14.578794	12793		
180453 NODE_11_length_16414_cov_13.733703	16446		
180454 NODE_386_length_3699_cov_13.323060	3731		
180455 NODE 215 length 395 cov 4.250633	427		

Click 'Submit'. A summary of the number of isolates and their lengths will be displayed. To confirm upload, click 'Upload'.

It is also possible to upload data for multiple isolates at the same time, but these must exist as single contigs for each isolate. To do this, select 'Read identifier from FASTA' in the isolate id field and select the field that you wish to use as the identifier in the 'identifier field', e.g. to use isolate names select 'isolate' here.



Provided the identifier used uniquely identifies the isolate you will get a confirmation screen. If the isolate name does not do this you'll probably have to use the database id number instead. Click 'Upload' to confirm.

PubM	LST / Users: Add   Q		s page (species)   Curato insert	ator's page (database)
Logged in: Ke	ith Jolley (keith). Log out	Change password		Toggle: i
Batch i	insert sequer	ices		
	ng sequences will be en			
BIGSdb id 938	Identifier field (isolate) 0021/93	Sequence length Co 465	omments Status Will upload	Upload
944	0030/93	465	Will upload	
946	0033/93	465	Will upload	

#### 6.12.2 Select from isolate query

As an alternative to selecting the isolate from a dropdown list (which can become unwieldy for large databases), it is also possible to upload sequence data following an isolate query.

Click the isolate query link from the curator's main page.

		e password			
tabase curator's	inte	rface	- Neisseri	ia PubN	ILST
🛃 Add, update or delet	e reco	rds			
Record type	Add	Batch Add	Update	or delete	Comments
users	+	++		?	
user groups	+	++		?	Users can be members of these groups - use for setting access permissions.
user group members	+	++		?	Add users to groups for setting access permissions.
curator permissions				?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users
					table.
isolates	+	++		owse   list   update	
isolate field extended attribute values	+	++		?	Add values for additional isolate field attributes.
projects	+	++		?	Set up projects to which isolates can belong.
project members	+	++		?	Add isolates to projects.
isolate aliases	+	++		?	Add alternative names for isolates.
PubMed links	+	++		?	
allele designations		++		?	Allele designations can be set within the isolate table functions.
sequences		++		?	The sequence bin holds sequence contigs from any source.
accession number links	+	++		?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++		?	Set up experiments to which sequences in the bin can belong.
experiment sequences				?	Add links associating sequences to experiments.
sequence tags		scan		?	Tag regions of sequences within the sequence bin with locus information.

Enter your search criteria. From the list of isolates displayed, click the 'Upload' link in the sequence bin column of the appropriate isolate record.

PubMLST / Users: Add   Que	es home   Curator's page (speci ry/update Jery/update   Batch insert	es)   Curator's page (database)			
Logged in: Keith Jolley (keith). 🔂 Log out   C	hange password	Help	🗹 🛛 Toggle: 🚯 🛛 Field help: id	1	<b>→</b> Go
Isolate query/update					
	ds		s		Modify
isolate 👻 =	▼ FAM18	+ 1 Order by: id			
		Display: 25	records per page ()		options
Reset Submit  1 record returned. Click the hyperlink fo Delete Tag scanning Delete ALL Scan	r detailed information. Projects Select project	▼ Link			
Delete Update Sequence New		Isolate fields 🛈	Seqbin	MLST	Finetyping antigens
belete opdate bin version	id isolate aliases cour	ntry year disease sp	ecies serogroup <sup>size (bp)</sup>	ST CIONAL P	PorA PorA FetA /R1 VR2 VR
delete update upload create	698 FAM18 NIBSC_3076; US Z4259		sseria C 2194961 ingitidis	1 11 ST-11 complex/ET-37 complex	5 2 F1-30 pdate update update

The same upload form as detailed above is shown. Instead of a dropdown list for isolate selection, however, the chosen isolate will be pre-selected.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Batch insert sequences	
This page allows you to upload sequence data for a specified isolate record in FASTA format.	
If an isolate id is chosen, then all sequences will be associated with that isolate. Alternatively, the isolate id, or any other iso rows of the FASTA file. This allows data for multiple isolates to be uploaded.	plate table field that uniquely defines the isolate, can be named in the identifier
Please note that you can reach this page for a specific isolate by querying isolates and then clicking 'Upload' within the isolate	ate table.
Please fill in the following fields - required fields are marked with an exclamation mark (!).	
- Paste in sequences in FASTA format:	Attributes isolate id: 1 698) FAM18 sender: ! method: run Id: assembly id: Options Don't insert sequences shorter than 250   bps. Link to experiment Alternatively upload FASTA file Action Select FASTA file: Reset Submit Browse No file selected.
Back	

### 6.12.3 Upload options

On the upload form, you can select to filter out short sequences from your contig list.

If your database has experiments defined (experiments are used for grouping sequences and can be used to filter the sequences used in *tag scanning*), you can also choose to upload your contigs as part of an experiment. To do this,

Database: Species home | Curator's page (species) | Curator's page (database) Users: Add | Query/update Isolates: Add | Query/update | Batch insert PubMLST Logged in: Keith Jolley (keith). Log out | Change password Toggle: i Batch insert sequences This page allows you to upload sequence data for a specified isolate record in FASTA format If an isolate id is chosen, then all sequences will be associated with that isolate. Alternatively, the isolate id, or any other isolate table field that uniquely defines the isolate, can be named in the identifier rows of the FASTA file. This allows data for multiple isolates to be uploaded. Please note that you can reach this page for a specific isolate by querying isolates and then clicking 'Upload' within the isolate table Please fill in the following fields - required fields are marked with an exclamation mark (!). Paste in sequences in FASTA format: - Attributes isolate id: ! Read identifier from FASTA \_ identifier field: id sender: ! • method: run id: assembly id: Options Don't insert sequences shorter than 250 bps Link to experiment: • Alternatively upload FASTA file -Action Select FASTA file: Reset Submit Browse... No file selected. Back

select the experiment from the dropdown list box.

### 6.13 Automated web-based sequence tagging

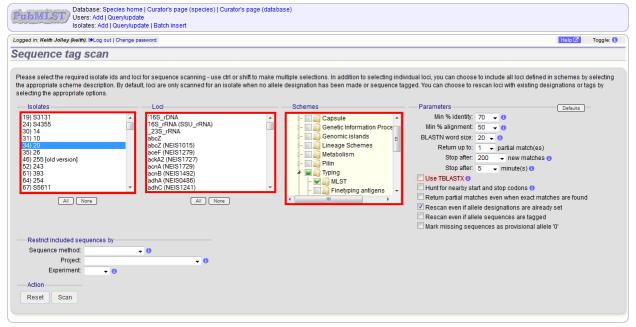
Sequence tagging, or tag-scanning, is the process of identifying alleles by scanning the sequence bin linked to an isolate record. Defined loci can either have a single reference sequence, that is defined in the locus table, or they can be linked to an external database that contains the sequences for known alleles. The tagging function uses BLAST to identify sequences and will tag the specific sequence region with locus information and an allele designation if a matching allele is identified by reference to an external database.

Select 'scan' sequence tags on the curator's index page.

tabaca auratar's int		assword		
	torf	200 -	Neisseria PubMLS	ST
	len	ace -	Neissena Fubille	
🔗 Add, update or delete re	corr	c		
		atch Add		Comments
40010	+	++	?	
abor groupo	+	++	?	Users can be members of these groups - use for setting access permissions.
abor group momotro	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table
	+	++	query   browse   list   batch update	
	+	++	?	Add values for additional isolate field attributes.
projecto	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags	S	can	?	Tag regions of sequences within the sequence bin with locus information.

Next, select the isolates whose sequences you wish to scan against. Multiple isolates can be selected by holding down the Ctrl key. All isolates can be selected by clicking the 'All' button under the isolate selection list.

Select either individual loci or schemes (collections of loci) to scan against. Again, multiple selections can be made.



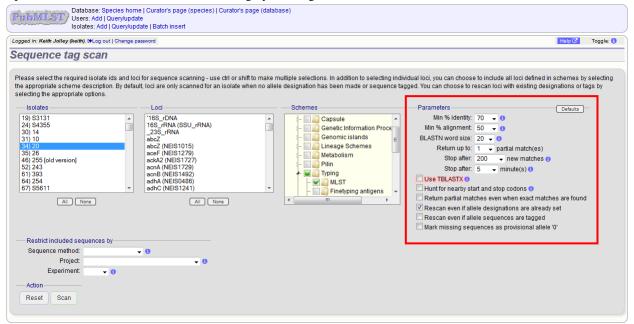
Choose your scan parameters. Lowering the value for BLASTN word size will increase the sensitivity of the search at the expense of time. Using TBLASTX is more sensitive but also much slower. TBLASTX can only be used to identify the sequence region rather than a specific allele (since it will only match the translated sequence and there may be multiple alleles that encode a particular peptide sequence).

By default, for each isolate only loci that have not had either an allele designation made or a sequence region scanned

will be scanned again. To rescan in these cases, select either or both the following:

- · Rescan even if allele designations are already set
- Rescan even if allele sequences are tagged

Options can be returned to their default setting by clicking the 'Defaults' button.



Press 'Scan'. The system takes approximately 1-2 seconds to identify each sequence (depending on machine speed and size of definitions databases). Any identified sequences will be listed in a table, with checkboxes indicating whether allele sequences or sequence regions are to be tagged.

quene	e tag	J SC	all												
olate Match	1 Locus	Allele	% identity	Alignment length	Allele lengt	1 E-value	Sequence bin id			rt Predicted end	Orientatior			Flag 🚯	
4) 20 exact		1	100.00	433	433	0.0	182791	7064 7496		7496 extract -	←	<b>v</b>			-
I) 20 exact		3	100.00	465	465	0.0	182750	1392 1856		1856 extract →	$\rightarrow$				-
4) 20 exact		1	100.00	490	490	0.0	182731	32577 33066		33066 extract →	←	<b>V</b>			-
		1	100.00	465	465	0.0	182815	19783 20247		20247 extract -	$\rightarrow$	<b>V</b>			-
4) 20 exact	gdh	1	100.00	501	501	0.0	182852	7516 8016	7516	8016 extract -	$\rightarrow$	<b>V</b>			-
I) 20 exact	pdhC	1	100.00	480	480	0.0	182819	13868 14347	13868	14347 extract -	$\rightarrow$	V			-
<li>4) 20 exact</li>	pgm	3	100.00	450	450	0.0	182831	24559 25008	24559	25008 extract ->	$\rightarrow$				-
Action Tag alleles	s/seque	nces										All None	All None		

Individual sequences can be extracted for inspection by clicking the 'extract  $\rightarrow$ ' link. The sequence (along with flanking regions) will be opened in another browser window or tab.

Checkboxes are enabled against any new sequence region or allele designation. You can also set a flag for a particular sequence to mark an attribute. These will be set automatically if these have been defined within the sequence definition database for an identified allele.

See also:

#### Sequence tag flags

Ensure any sequences you want to tag are selected, then press 'Tag alleles/sequences'.

If any new alleles are found, a link at the bottom will display these in a format suitable for automatic allele assignment by *batch uploading to sequence definition* database.

#### See also:

Offline curation tools

Automated offline sequence tagging

### 6.14 Projects

#### 6.14.1 Creating the project

The first step in grouping by project is to set up a project.

Click the add (+) project link on the curator's main page.

Add, update or delete r	tei	rface -		
		acc -	Neisseria PubMLS	ST
Add undate or delete r				
Add, update of delete i	eco	rds		
Record type	Add	Batch Add	Update or delete	Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table.
isolates	+	++	query   browse   list   batch update	
late field extended attribute values	+	++	?	Add values for additional isolate field attributes.
projects	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
isolate aliases	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags		scan	?	Tag regions of sequences within the sequence bin with locus information.
Database configuration				
_				
			pdate or delete	Comments
loci +		++	?	
da		nk scan		
		++	? Add alternative name	es for loci. These can also be set when you batch add loci.
locus aliases +				
locus aliases + PCR reactions +		++	Set up in silico PCR r	reactions. These can be used to filter genomes for tagging to specific repetitive loci.

Enter a short description for the project. This is used in drop-down list boxes within the query interfaces, so make sure it is not too long.

You can also enter a full description. If this is added, the project description can displayed at the top of an isolate information page (but see 'isolate\_display' flag below). The full description can include HTML formatting, including image links.

There are additionally two flags that affect how projects are listed:

- isolate\_display Setting this is required for the project and its description to be listed at the top of an isolate record (default: false).
- list Setting this is required for the project to be listed in a page of projects linked from the main contents page.

#### Click 'Submit'. Database: Species home | Curator's page (species) | Curator's page (database) PubMLST Users: Add | Query/update Isolates: Add | Query/update | Batch insert Logged in: Keith Jolley (keith). Hog out | Change password Toggle: 🚯 Add new project description Please fill in the fields below - required fields are marked with an exclamation mark (!). Record Action id:! 3 ÷ Reset Submit short description:! MRF Meningococcus Genome Library 0 curator:! Keith Jolley (keith) datestamp:! 2015-07-02 н Meningococcus Genome Library</a> is a collaboration between Public Health England, the Scottish Haemophilus, Legionella, Meningococcus and Pneumococcus Reference (SHIMPR) Laboratory, The Wellcome Trust Sanger Institute and the University of Oxford, funded by the Meningitis Research ÷ ิด isolate display: 💿 true 🔘 false 🜖 list 💿 true 🔘 false 🜖

### 6.14.2 Explicitly adding isolates to a project

Explicitly adding isolates to the project can be done individually or in batch mode. To add individually, click the add (+) project member link on the curator's main page.

iged in: Keith Jolley (keith). @Log ou Atabase curator's	it   Cha	vide bassword		
atabase curator's				
	int	erface -	- Neisseria	PubMLST
🎻 Add, update or delet	te re	cords		
	_	Detab		
Record type	Add	Batch Add	Update or dele	te Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in th
				users table.
isolates	+	++	query   browse	list
			batch update	
isolate field extended attribute values	+	++	?	Add values for additional isolate field attributes.
projects	+	++	2	Cature ansists to which inclutes and below
project members	+	++	2	Set up projects to which isolates can belong. Add isolates to projects.
isolate aliases	<u> </u>	++	2	Add isolates to projects.
PubMed links	+	++	· · · · · · · · · · · · · · · · · · ·	Add alternative names for isolates.
		++	2	Allele designations and he activities the index to be for the sectors
allele designations sequences		++	2	Allele designations can be set within the isolate table functions.
accession number links	+	++	2	The sequence bin holds sequence contigs from any source.
	+		2	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences				Add links associating sequences to experiments.
sequence tags		scan	?	Tag regions of sequences within the sequence bin with locus information.
🎽 Database configurati	on			
Table	Add	Batch Add	Update or delete	Comments
loci	+	H+	?	
100		abank scan	r.	
locus aliases	uata +	++	2 Add	alternative names for loci. These can also be set when you batch add loci.
locus allases	+	++		•
BCD reactions	+	++	Set	up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
PCR reactions				
PCR reactions nucleotide probes isolate field extended attributes	+	++		ne nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.

Select the project from the dropdown list box and enter the id of the isolate that you wish to add to the project. Click 'Submit'.

PubliALST Isolates: Add   Querylupdate Isolates: Add   Querylupdate   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Add new project member	
Please fill in the fields below - required fields are marked with an exclamation mark (!).  Record  Project id:! MRF Meningococcus Genome Library  isolate id:! 18968  curator:! Keith Jolley (keith) datestamp:! 2014-07-10	

To add isolates in batch mode. Click the batch add (++) project members link on the curator's main page.

ed in: Keith Jolley (keith). @Log ou tabase curator's		ige password		
tabase curator's				
	Inte	erface -	Neisser	a PubMLST
Add, update or delet	te rec	ords		
Record type	Add	Batch Add	Update or	delete Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in t users table.
isolates	+	++	query   brow batch up	
solate field extended attribute values	+	++	?	Add values for additional isolate field attributes.
projects	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
isolate aliases	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags		scan	?	Tag regions of sequences within the sequence bin with locus information.
Database configurati Table		Dotob Add H	udate or delate	Commonto
loci	Add E	Batch Add U	pdate or delete ?	Comments
1001		++ bank scan	f	
locus aliases	4	++	2	Add alternative sames for lasi. These can also be set when you batch add lasi
PCR reactions	+	++		Add alternative names for loci. These can also be set when you batch add loci.
				Set up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		Define nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.

Download an Excel submission template:

PTITIMLST Users: Add   Query/update Isolates: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). Log out   Change password	Toggle: i
Batch insert project members	
This page allows you to upload project member data as tab-delimited text or copied from a spreadsheet.  • Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.  • Download tab-delimited header for your spreadsheet - use Paste special → text to paste the data.  • Download submission template (xlsx format)	
Paste in tab-delimited text (include a field header line).	- Action
	Reset Submit
Back	

You will need to know the id number of the project - this is the id that was used when you created the project. Fill in the spreadsheet, listing the project and isolate ids. Copy and paste this to the web upload form. Press 'Submit'.

ged in: Keith Jolley (ke	ith). Log out   Change password						Toggle: i
atch insert	project members						
iis page allows you t	o upload project member data as tal	b-delimited text or copied from a s	spreadsheet.				
<ul> <li>Field header na</li> </ul>	mes must be included and fields ca	an be in any order. Optional fields	can be omitted if you wish.				
	lelimited header for your spreadshe nission template (xlsx format)	eet - use Paste special → text to p	aste the data.				
-Paste in tab-delim	ted text (include a field header line)	)		 Action -		_	
project_id 3 18968 3 18969 3 19023 3 19024 3 19025 3 19026 3 19027 3 19028 3 19029 3 19031 3 19031 3 19958 3 19959	isolate_id			Reset	Submit		

# 6.15 Isolate record versioning

Versioning enables multiple versions of genomes to be uploaded to the database and be analysed separately. When a new version is created, a copy of the provenance metadata, and publication links are created in a new isolate record. The sequence bin and allele designations are not copied.

By default, old versions of the record are not returned from queries. Most query pages have a checkbox to 'Include old record versions' to override this.

Links to different versions are displayed within an isolate record:

(PubMLST) B	tuery: Search   Browse   Profile/S reakdown: Isolate fields   Schen inks: Contents   Home   Options		ase submissions		
Full informat	ion on isolate 5				
Provenance/meta	adata				
id:	21	source:	throat swab	update history:	1 update show details
isolate:	5	epidemiology:	endemic	date entered:	2014-06-30
strain designation:	NG: P1.ND,ND: F-ND: ST-ND (-	-) species:	Neisseria meningitidis	datestamp:	2014-06-30
country:	Germany	serogroup:	NG		
continent:	Europe	comments:	Bavarian carriage study		
region:	Volkach	sender:	Ulrich Vogel, Univerity of Wuerzburg, Wuerzburg,		
year:	1999		Germany		
age yr:	18	curator:	Keith Jolley, University of Oxford, UK (E-mail: keith.jolley@zoo.ox.ac.uk)		
disease:	carrier		keiur.joney@200.0x.ac.uk)		
Versions					
More than one version	of this isolate record exist.				
Older versions:	20				
Publication (1)					
Claus H, Maide	en MC, Wilson DJ, McCarthy ND,	Jolley KA, Urwin R, Hessler F, Frosch M	, Vogel U (2005). Genetic analysis of meningococci o	arried by children a	nd young adults. J Infect Dis 191:1263-71
822 isolates					
Schemes and loc	i				
└ 🛄 No loci availab	le for analysis.				

The different versions will also be listed in analysis plugins, with old versions identified with an [old version] designation after their name.

To create a new version of an isolate record, query or browse for the isolate:

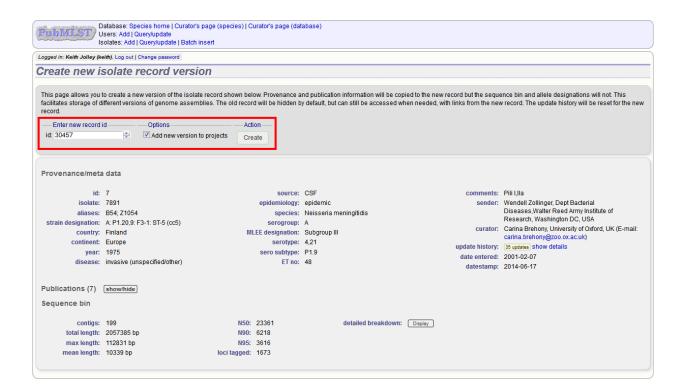
ged in: Keith Jolley (keith). 🕒 Log ou	t I Chano	e password		
			M	
atabase curator's	Inte	rface -	Neisseria	PUDMLST
💰 Add, update or delet	te reco	ords		
Record type	Add	Batch Add	Update or dele	te Comments
users	+	++	?	
user groups	+	++	?	Users can be members of these groups - use for setting access permissions.
user group members	+	++	?	Add users to groups for setting access permissions.
curator permissions			?	Set curator permissions for individual users - these are only active for users with a status of 'curator' in th
				users table.
isolates	+	++	query   browse     batch update	
isolate field extended attribute values	+	++	?	Add values for additional isolate field attributes.
projects	+	++	?	Set up projects to which isolates can belong.
project members	+	++	?	Add isolates to projects.
isolate aliases	+	++	?	Add alternative names for isolates.
PubMed links	+	++	?	
allele designations		++	?	Allele designations can be set within the isolate table functions.
sequences		++	?	The sequence bin holds sequence contigs from any source.
accession number links	+	++	?	Associate sequences with Genbank/EMBL accession number.
experiments	+	++	?	Set up experiments to which sequences in the bin can belong.
experiment sequences			?	Add links associating sequences to experiments.
sequence tags	S	can	?	Tag regions of sequences within the sequence bin with locus information.
<b>6</b>				
📂 Database configurati	on			
Table	Add Ba	atch Add II	pdate or delete	Comments
loci	+	++	2	Comments
1001		nk scan		
locus aliases	+	++	? Add	alternative names for loci. These can also be set when you batch add loci.
PCR reactions	+	++		up in silico PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci.
nucleotide probes	+	++		ne nucleotide probes for in silico hybridization reaction to filter genomes for tagging to specific repetitive loci.
isolate field extended attributes	+	++		the additional attributes to associate with values of a particular isolate record field.

Click the 'create' new version link next to the isolate record:

Publ	<b>TLST</b>	Users: Ad	: Species hor d   Query/upd \dd   Query/up	ate			Curator's page	(database)										
Logged in:	Keith Jolle	ey (keith). 🗭 Log	g out   Change	passw	vord									Toggl	e: 🜖 Field help: id			<b>-</b> Co
Brow	se Ne	eisseria	PubMl	LS	T datab	ase												
O Di (	Display:	d ascending	ds per page	0		•	Action Browse all re	cords										
Dele	te te ALL	Tag scar Scan	Se	rojec lect p			nformation. Link											
Delete	Update S	Sequence bin	New version	n id	isolate	-	iases		late fi vear	elds 🕦	disease	species		ST	MLST clonal complex	Fine PorA VR1	typing anti	
delete	update	upload	create	10			2803; Z1001					Neisseria meningitidis	serogroup	4	ST-4 complex/subgroup IV	5-2 update		F1-5 update
	update	upload	create	2			2822: Z1035					Neisseria meningitidis		1	ST-1 complex/subgroup I/I	5-2 update		F5-1 update
	update	upload	create	3	M00242905			UK	2000	invasive (	unspecified/other)	Neisseria meningitidis	в	1099		19 update	15 update	add
delete	update	upload	create	4	M1027	B43; NIBSO	C_3076; Z1043	USA	1937	invasive (	unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex/subgroup IV	add	add	add
delete	update	upload	create	5	M00240227			UK	2000	invasive (	unspecified/other)	Neisseria meningitidis	В	1100	ST-32 complex/ET-5 complex	7 update	16 update	add
delete	update	upload	create	6	M00282207			UK	2000	invasive (	unspecified/other)	Neisseria meningitidis	W	1101	ST-22 complex	add	add	add
delete	update	upload	create	7	7891	B54; NIBS0	C_2760; Z1054		1975	invasive (	unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex/subgroup III	20 update	9 update	F3-1 update
delete	update	upload	create	8	M00242007							Neisseria meningitidis		1102	ST-18 complex	add	14 update	add
delete	update	upload	create	9	0021/84			Czech Republic	1984	invasive (	unspecified/other)	Neisseria meningitidis	W	114	ST-22 complex	add	add	add
doloto	undata	unload	croate	10	6740	D72 NIDO	0704-71070	Conodo	1071	invocivo (r	unepacified/other)	Noiceoria moningitidir	Α	4	QT 1 complex/cubaroup I/II	10.1	2	EE 1 madete

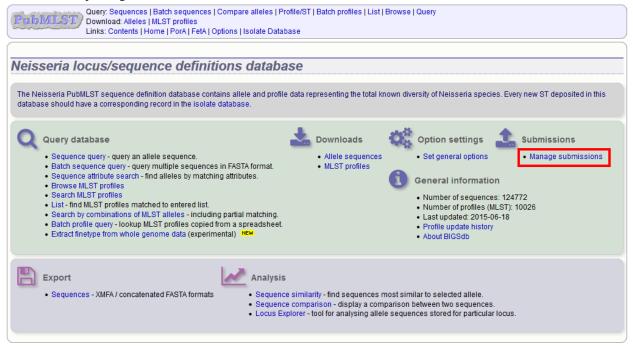
The isolate record will be displayed. The suggested id number for the new record will be displayed - you can change this. By default, the new record will also be added to any projects that the old record is a member of. Uncheck the 'Add new version to projects' checkbox to prevent this.

Click the 'Create' button.



## Curating submitted data

Data may be submitted by users using the automated submission system if it has been enabled for a specific database. As a curator, you will be notified of pending submissions when you log in to the curator's interface or if you access the 'Manage submissions' links from the standard contents page. Additionally, if your user account has the 'submission\_emails' flag set in the users' table you will also receive E-mail notification of new submissions for which you have sufficient privileges to curate.



Any submissions for which you have sufficient privileges to curate will be shown.

ogged in: Keith Jolley (keith). 🗭 Log	out I Change password	
lanage submissi		
Submit new data		
Data submitted here will go in to	a queue for handling by a curator or by an automated scrip	pt. You will be able to track the status of any submission.
Submission type:		
<ul><li>alleles</li><li>MLST profiles</li></ul>		
New allele sequence sub	missions waiting for curation	
Your account is authorized to ha	ndle the following submissions:	
Submission id	Submitted Updated Submitter Locus	Sequences
	862_84622 2015-06-23 2015-06-23 Joe Bloggs NEIS000	
Return to index page		

# 7.1 Alleles

Click the link to the appropriate submission on the 'Manage submissions' page.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Logged in: Keith Jolley (keith). @Log out   Change password Help C	
Manage submissions	
Submit new data	
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.	
Submission type:	
alleles     MLST profiles	
New allele sequence submissions waiting for curation	
Your account is authorized to handle the following submissions:	
Submission id         Submitted         Updated         Submitter         Locus         Sequences           BIGSdb_20150709132553_20864_89729         2015-07-09         2015-07-09         Joe Bloggs         NEIS0001 (lpxC)         3	
Return to index page	

You will see a summary section that describes details about how the sequences were obtained. There should also be link here to download all the sequences in FASTA format.

ged in: Keith Jolley (keith	). DeLog out   Change password							He
urate submis	sion							
	b_20150709132553_20864_897	1						
- Summary		Sequence Identifier			Sequence	Complete CDS	Status	Assigned allele
	alleles	UK322	924		CCGC ACATCAGAAATTGTGGATAA		pending +	Curate
	Joe Bloggs, University of Oxford, UK	UK323	924	ATGCTGCAAAGAACTT	TGGC ACATCAGAACTTGTGGATAA	~	pending +	Curate
datestamp:		UK347	924	ATGCTGCAAAGAACTT	TGGC ACATCAGAAATTGTGGATAA	<b>~</b>	pending 👻	Curate
status:	pending NEIS0001 (lpxC)	Batch cui	rate					Update
sequences:					Analysis			
		Message	s		Archive	andia a Glassi		
technology:					Archive of submission and any supp	forting mes.		
read length: coverage:					Download TAR			
assembly:		Message:	4000	nd Send now				
	ue novo	messaye.	Apper	Send now				

There will also be a table summarizing the sequences in the submission and their current submission status.

PubMLST / Do	ery: Sequences   Batch sequences   Co wnload: Alleles   MLST profiles ks: Contents   Home   PorA   FetA   Optio		file/ST   Batch profiles   List   Browse   Query ase				
Logged in: Keith Jolley (keit	h). HLog out   Change password					Hel	p 🕑
Curate submi	ssion						
— Summary— type: submitter: datestamp: status:	pending	29 Sequences Identifier Length UK322 924 UK323 924 UK347 924 Batch curate	Sequence АГССТССАЛАСААСТТСЕGC АСАГСАБААТТБТБСАТАА АГССТБСАЛАБААСТТБЕGC АСАГСАБААСТТБТБСАТАА АГБСТБСАЛАБААСТТБЕGC АСАГСАБААЛТБТБСБАТАА	Complete CDS CDS	Status pending + pending +	Curate	
iocus: sequences: technology: read length: coverage: assembly assembly software:	llumina 100-199 20-49x de novo	Messages Message: Appe	Archive Archive of submission and any supp Download R nd Send now	orting files:			

#### 7.1.1 Individual allele curation

Individual sequences can be curated singly by clicking the 'Curate' links next to the sequence in the table. If you have supporting data attached to the submission, e.g. Sanger trace files then you may need to assess the submission based on the policy of the database.

						H
urate submission						
ubmission: BIGSdb 20150709132553 20864 897	29					
	Identifier Ler	nath	Sequence	Complete CDS	Status	Assigned allele
type: alleles		-	ICGGC ACATCAGAAATTGTGGATAA	~	pending 👻	Curate
submitter: Joe Bloggs, University of Oxford, UK	UK323 9	24 ATGCTGCAAAGAACT	TTGGC ACATCAGAACTTGTGGATAA	<b>~</b>	pending +	Curate
datestamp: 2015-07-09	UK347 9	24 ATGCTGCAAAGAACT	TTGGC ACATCAGAAATTGTGGATAA	~	pending -	Curate
status: pending	Batch curate	e				Update
locus: NEIS0001 (lpxC)						opullo
sequences: 3 PAS	Messages		Archive			
technology: Illumina			Archive of submission and any supp	orting files:		
read length: 100-199						
coverage: 20-49x			Download TAR			
assembly: de novo	Message: A	ppend Send now				
assembly: de novo						

Clicking this link takes you to the curation interface *single sequence upload page*. The upload form will be filled with details from the submission. You may wish to manually change the status from the dropdown list of values.

PubMLST Use	abase: Species home   Curator's page (species)   Curator's page (database) ers: Add   Query/update d: Add ST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith	h). @Log out   Change password Help 🖸	Toggle: 🚯
Add new allele	e sequence	
	elow - required fields are marked with an exclamation mark (!).	
	NEIS0001 -	
allele id:!		
sequence.	ATGCTGCAAAGAACTTCGGCGAAATCCATCAGCGTTACCGGAGTCGCCTGCATTCCGGCGAACGGGTCGCACTG ACCCTGCACCCCGCGCCTGAAAACAGCGGGATTTCCTTCC	
status:!	unchecked -	
sender:!	Bloggs, Joe (jbloggs)	
curator:	Keith Jolley (keith)	
date entered:	2015-07-09	
datestamp:!	2015-07-09	
comments:		.A
Flags:	atypical contains IS element downstream fusion	

Clicking 'Submit' from this form will define the new allele and add it to the database. A link on the confirmation page will take you back to the submission management page.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert		
Logged in: Keith Jolley (keith). @Log out   Change password	Help 🖉	Toggle: 🚺
Add new allele sequence		
Sequence NEIS0001 (lpxC): 210 added! Add another Return to submission Back to main page		

You will find that the status of the newly assigned sequence has changed in the summary table. The assigned value and status are determined on display and should always reflect the live database values.

PubMLST Do	uery: Sequences   Batch sequences   Co ownload: Alleles   MLST profiles nks: Contents   Home   PorA   FetA   Opti	· · · ·	ofile/ST   Batch profiles   List   Browse   Query ase			
Logged in: Keith Jolley (kei	th). @Log out   Change password					Help 🖉
Curate submi	ission					
— Summary — type: submitter: datestamp: status:	lllumina 100-199 20-49x de novo	29 Sequences Identifier Lengt UK322 924 UK323 924 UK347 924 Batch curate Messages Message: Appe	ATGCTGCAAAGAACTTCGGC ACATCAGAAATTGTGGATAA ATGCTGCAAAGAACTTGGC ACATCAGAACTTGTGGATAA ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA Archive Archive Download	· · ·	Status assigned pending ↓ pending ↓	Assigned allele 210 Crute Curate Update

### 7.1.2 Batch allele curation

Often, you will want to batch upload submitted sequences. This can be done by clicking the 'Batch curate' button.

**Note:** Batch curation is only available for loci that do not have extended attributes defined. Entries for these loci require additional values set for these additional fields and so need to be handled individually.

	th). Hog out   Change password						He
urate submi	ssion						
ubmission: BIGS	1b 20150709132553 20864 897	29					
			es —				
		Identifier	Lenath	Sequence	Complete CDS	Status	Assigned allele
	alleles	UK322	924	ATGCTGCAAAGAACTTCGGC ACATCAGAAATTGTGGATAA	<b>~</b>	assigned	210
	Joe Bloggs, University of Oxford, UK	UK323	924	ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA	<ul> <li>Image: A set of the set of the</li></ul>	pending 👻	Curate
	2015-07-09	UK347	924	ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA	<ul> <li>Image: A set of the set of the</li></ul>	pending 👻	Curate
	pending	Batch cur	rate				Update
	NEIS0001 (IpxC)						
sequences:	3 FAS	Message	es	Archive			
technology:	Illumina			Archive of submission and any supp	orting files:		
read length:	100-199						
coverage:	20-49x						
assembly:	de novo	Message:	Appe	nd Send now			

This takes you to the batch FASTA upload page in the curators' interface.

The upload form will be filled with details from the submission. You may wish to manually change the status from the dropdown list of values.

PuliML	Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert		
Logged in: Keith	Jolley (keith). @Log out   Change password	Help 🗗	Toggle: 🚯
Batch in	sert sequences		
only). Do not i	ows you to upload allele sequence data in FASTA format. The identifiers in the FASTA file will be used unless you select the option to use the next available id nclude the locus name in the identifier in the FASTA file. nat you can not use this page to upload sequences for loci with extended attributes.	l (loci with in	teger ids
	NEIS0001 (IpxC)		
	Blogas, Joe (jblogas)		
sequence (FASTA):!	>UK323 AIGCIGCAAAGAACTTIGGCGAAATCGATCGGCGTACCGGGGTCGCATTCGGGG GAACGCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG		
ignored.	all sequences that are not complete reading frames - these must have a start and in-frame stop codon at the ends and no internal stop codons. Existing sec e sequence similarity check xt available id (only for loci with integer ids) 	quences are	also

Click 'Check' on this form will perform some standard checks before allowing you to upload the sequences.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert										
Logged in: Keith Jolley	(keith). 🕒 Log out   Change pa	assword		E	Help 🖉	Toggle: 🚯				
Batch inser	t sequences									
Sequence chec Original designatio UK323 UK347	n Allele id Status	Action Upload valid sequences								

A link on the confirmation page will take you back to the submission management page.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). [9Log out   Change password Help 2	Toggle: 🚯
Batch insert sequences	
Upload succeeded. Return to submission Upload more   Back to main page	

The status of the sequences should reflect their newly assigned status.

PrifiMILST Download: A Links: Conte	ences   Batch sequences   Compare alleles   Pri lleles   MLST profiles nts   Home   PorA   FetA   Options   Isolate Datab				
Logged in: Keith Jolley (keith). Hog o	ut   Change password				Help 🖉
Curate submission					
Submission: BIGSdb_2015	0709132553_20864_89729				
Guinnary	Identifier Lengt	Sequence	Complete CDC	Status Assigned allele	
type: alleles submitter: Joe Blog datestamp: 2015-07 status: pending	UK322 924 UK323 924 UK323 924 UK347 924	АТВСТБСАЛАВЛАСТТСВОС АСАТСАВАЛАТТБТВБАТАЛ АГВСТБСАЛАВЛАСТТБВС АСАТСАВАЛТТБТВБАТАЛ АГВСТБСАЛАВЛАСТТБВС АСАТСАВАЛТТБТВБАТАЛ АГСНИС АССНИС	*	assigned 210 assigned 211 assigned 212 Action	
locus: NEIS000 sequences: 3 rAS technology: Illumina	11 (lpxC)	Archive of submission and any supp	orting files:	Close submission	
read length: 100-199 coverage: 20-49x assembly: de novo assembly software: Velvet	Message: App	nd Send now			

#### 7.1.3 Rejecting sequences

Sometimes you may need to reject all, or some of, the sequences in a submission. You can do this by changing the value in the status dropdown box next to each sequence. Click 'Update' to make the change.

igea in: Keith Jolley (Keith	<ol> <li>Hog out   Change password</li> </ol>								Help
urate submis	ssion								
ubmission: BIGSd	b_20150709132553_20864_897	29							
- Summary		Sequence	ces						
typo:	alleles	Identifier				Complete CDS	Status	Assigned allele	
	Joe Bloggs, University of Oxford, UK	UK322			ICGGC ACATCAGAAATTGTGGATAA	<ul> <li>✓</li> </ul>	rejected 👻	Curate	
datestamp:		UK323			ITGGC ACATCAGAACTTGTGGATAA	<ul> <li>✓</li> </ul>	rejected 👻	Curate	
	pending	UK347	924	ATGCTGCAAAGAACTI	ITGGC ACATCAGAAATTGTGGATAA	<ul> <li>✓</li> </ul>	rejected 👻	Curate 3	
	accepted - data uploaded	Batch cu	rate					Update	
	NEIS0001 (lpxC)				Anakin				
sequences:		Messag	as		Archive				
					Archive of submission and any supp	orting files:			
technology:					Download TAR				
read length:									
coverage:		Message:	Apper	nd Send now					
assembly:	de novo								
assembly software:									

### 7.1.4 Requesting additonal information

You can send a message to the submitter by entering it in the Messages box and clicking 'Send now'. This will append a message to the submission and send an update to the submitter so that they can respond.

#### 7.1.5 Closing the submission

You can add a message to the submitter by entering it in the message box and clicking 'Append'. Once sequences have all been either assigned or rejected, the 'Close submission' button will be displayed. Click this to close the submission. The submitter will be notified of their submission status.

PubMLST Do	ery: Sequences   Batch sequences   Co wnload: Alleles   MLST profiles iks: Contents   Home   PorA   FetA   Optic			s   List   Browse	Query				
Logged in: Keith Jolley (kei	th). @Log out   Change password								Help 🗗
Curate submi	ssion								
Submission: BIGS	ib_20150709132553_20864_8972	29							
Summary		Sequences							
fyner	alleles	Identifier Length		Sequence		Complete CDS	Status	Assigned allele	
	Joe Bloggs, University of Oxford, UK				CAGAAATTGTGGATAA	~	rejected 👻		
	2015-07-09				CAGAACTTGTGGATAA	~	rejected 👻		
	pending	UK347 924	ATGCTGCAAAGAACT	TIGGC ACAI	CAGAAATTGTGGATAA	~	rejected 👻		
	NEIS0001 (IpxC)							Update	
sequences:		Messages —				Arch	ive		
technology:		Timestamp	User		Message			on and any suppo	rtina files:
read length:					s are not NEIS0001 a	lleles	_		
coverage:						Downl	oad TAR		
assembly:									
assembly software:	Velvet					Actio	n		
				Message:	Append Send	Clos	e submissio	n	

# 7.2 Profiles

Click the appropriate submission on the 'Manage submissions' page.

Ouery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   Pork   Fetk   Options   Isolate Database
Logged in: Keith Jolley (keith). (HLog out   Change password Help C
Manage submissions
Submit new data
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.
Submission type:
alleles     MLST profiles
New allelic profile submissions waiting for curation
Your account is authorized to handle the following submissions:
Submission id         Submitted         Updated         Submitter         Scheme         Profiles           BIGSdb_20150709134405_4219_35535         2015-07-09         2015-07-09         Joe Bloggs         MLST         3
Return to index page

You will see a table summarizing the profiles in the submission and their current status.

Query: Sequences   Batch sequences Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA	Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Dptions   Isolate Database	
Logged in: Keith Jolley (keith). House out   Change password		Help 🗗
Curate submission		
Submission: BIGSdb_20150709134405_4219_34	Identifier         adk         abcZ         aroE   fumC         gdh         pdhC         pgm         Status         Assign           UK32         43         2         12         32         3         2         pending         Image: Comparison of the c	urate

### 7.2.1 Individual profile curation

Individual profiles can be curated singly by clicking the 'Curate' links next to the profile in the table.

Query: Sequences   Batch sequences   Co Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Opti					atch p	rofiles	s   List	Brov	vse   Quer	у			
Logged in: Keith Jolley (keith). DLog out   Change password													Help 🗹
Curate submission													
Submission: BIGSdb_20150709134405_4219_3553	5 — Profiles												
ounimary	Identifier		ahc7	aroE	fumC	adh	ndhC	nam	Status		Assigned ST		
type: profiles	UK32	43	2	12	32	32	3	2	pending		Curate		
submitter: Joe Bloggs, University of Oxford, UK	UK33	7	56	4	3	2	12	12	pending		Curate 3		
datestamp: 2015-07-09	UK34	76	3	5	3	87	43	34	pending	-	Curate 3		
status: pending	Batch c	urate									Update		
Messages Archive Archive Archive of submissi	on and any	suppo	orting	files:									

Clicking this link takes you to the curation interface *single profile upload page*. The upload form will be filled with details from the submission.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert												
Logged in: Keith Jolley (ke	ith). ⊕Log out   Change passw	ord		Help 🗹	Toggle: 🚺							
Add new MLS	ST profile											
~												
Please fill in the fields	below - required fields are	marked with an exclamation	mark (!).									
	10056											
adk: !												
abcZ: !												
aroE: !	12											
fumC: !	32 🜲											
gdh: !	32											
pdhC: !	3											
pgm: !	2											
sender: !	Bloggs, Joe (jbloggs)	<b>~</b>										
clonal_complex:												
	Keith Jolley (keith)											
date_entered: !												
datestamp: !	2015-07-09											
PubMed ids:												
- Action												
Reset Subn	nit											

Clicking 'Submit' from this form will define the new profile and add it to the database. A link on the confirmation page will take you back to the submission management page.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Loci: Add MLST profiles: Add   Query/update   Batch insert		
Logged in: Keith Jolley (keith). @Log out   Change password	Help 🖓	Toggle: 🚯
Add new MLST profile		
ST-10056 added!		
Return to submission Add another   Back to main page		

You will find that the status of the newly assigned profile has changed in the summary table. The assigned value and status are determined on display and should always reflect the live database values.

Ouery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database												
Logged in: Keith Jolley (keith). 🗈 Log out   Change password												
Curate submission												
Submission: BIGSdb_20150709134405_4219_3553 — Summary	5 Profiles Identifier UK32 UK33 UK34 Batch c	adk 43 7 76	abcZ 2 56 3	aroE 12 4 5	fumC 32 3 3	gdh 32 2 87	pdhC 3 12 43	pgm 2 12 34	Status assigne pending pending	ed 👻	Assigned ST 10056 Curate Curate Update	-
Messages Archive Archi	on and any	suppo	orting	files:								

#### 7.2.2 Batch profile curation

Often, you will want to batch upload submitted profiles. This can be done by clicking the 'Batch curate' button.

Query:         Sequences         Batch sequences         Compare alleles         Profile/ST         Batch profiles         List         Browse         Query           Download:         Alleles         MLST profiles         Download:         Alleles         MLST profiles           Links:         Contents         Home         PorA         FetA         Options         Isolate         Database													
Logged in: Keith Jolley (keith). DeLog out   Change password													
Curate submission													
— Summary type: submitter: datestamp:	Ib_20150709134405_4219_3553 profiles Joe Bloggs, University of Oxford, UK 2015-07-09 pending Archive	5 Profiles Identifier UK32 UK33 UK34 Batch c	adk 43 7 76	abcZ 2 56 3	aroE 12 4 5	fumC 32 3 3	gdh 32 2 87	pdhC 3 12 43	pgm 2 12 34	Status assigned pending pending	<ul> <li>Curate</li> </ul>		
Message: Append	Archive of submiss Download	ion and any	suppo	orting	files:								

This takes you to the *batch profile upload page* in the curators' interface.

#### The upload form will be filled with details from the submission.

PubM		sers: Add   0 oci: Add	Query/updat	te	; page (spec Batch insert	ies)   Curator's page (c	database)		
Logged in: K	(eith Jolley (ke	ith). ເ⊮Log ou	it   Change pa	assword				Help 🖾 Toggle:	0
Batch	insert	MLST	profile	S					
• Fie • You it fo • Do	eld header na u can choose or each profil	ames must e whether o e record. delimited he	be included r not to inclu eader for you	l and fields Ide a ST fiel ur spreadsh	can be in an d - if it is om			h. ically. If however, you include it in the header line, then you must also provi	ide
	e paste in ta				ader line) —			Parameters	
adk 7 76	abcZ 56 3	aroE 4 5	fumC 3 3	gdh 2 87	pdhC 12 43	pgm 12 34		Sender: Bloggs. Joe (jbloggs) Value will be overridden if you include a sender field in your pasted data. Ignore duplicate profiles Action Reset Submit	
Back									

After upload, a link on the confirmation page leads back to the submission management page.

Database: Species home   Curator's page (species)   Curator's page (database)           Users: Add   Query/update           Loci: Add           MLST profiles: Add   Query/update   Batch insert		
Logged in: Keith Jolley (keith). DeLog out   Change password	Help 🗹	Toggle: 🚺
Batch insert MLST profiles		
Database updated ok		
Return to submission Back to main page		

The status of the profiles should reflect their newly assigned status.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database														
Logged in: Keith Jolley (keit	Logged in: Keith Jolley (keith). @Log out   Change password Help 2													
Curate submi	ssion													
— Summary type: submitter: datestamp:	profiles Joe Bloggs, University of Oxford, UK 2015-07-09 pending	5 Profiles Identifier UK32 UK33 UK34 Message	adk 43 7 76 ges	abcZ 2 56 3	12 4 5	fumC 32 3 3	32 2 87	3 12 43 — A Arct Dov	2 12 34 rchive nive o vnloa	assigned assigned assigned	<b>Issigned ST</b> 10056 10057 10058 n and any sup			

### 7.2.3 Rejecting profiles

Sometimes you may need to reject all, or some of, the profiles in the submission. This may be because isolate data had not been made available, against the policy of the database. You can do this by changing the value in the status dropdown box next to each profile. Click 'Update' to make the change.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database														
Logged in: Keith Jolley (keit	h). ⊕Log out   Change password													Help 🖉
Curate submission														
— Summary type: submitter: datestamp: status:	b_20150709134405_4219_35533 profiles Joe Bloggs, University of Oxford, UK 2015-07-09 pending accepted - data uploaded	5 Profiles Identifier UK32 UK33 UK34 Batch cr	adk 43 7 76	abcZ 2 56 3	aroE 12 4 5	fumC 32 3 3	gdh 32 2 87	pdhC 3 12 43	pgm 2 12 34	Status rejected rejected rejected	• •	Assigned ST Curate Curate Curate Update		
		Messag	jes —						rchive		on an	nd any suppor	tina files:	
		Message:	An	pend	Se	end no	i			d TAR				
				pond		, na na								

#### 7.2.4 Requesting additonal information

You can send a message to the submitter by entering it in the Messages box and clicking 'Send now'. This will append a message to the submission and send an update to the submitter so that they can respond.

#### 7.2.5 Closing the submission

You can add a message to the submitter by entering it in the message box and clicking 'Append'. Once profiles have all been either assigned or rejected, the 'Close submission' button will be displayed. Click this to close the submission. The submitter will be notified of their submission status.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query         Download: Alleles   MLST profiles         Links: Contents   Home   PorA   FetA   Options   Isolate Database													
Logged in: Keith Jolley (keith). HLog out   Change password										Help 🗹			
Curate submission													
Submission: BIGSdb 20150709134405 4219 3553	5												
Summary	Profiles												
type: profiles	Identifier		_										
submitter: Joe Bloggs, University of Oxford, UK	UK32	43	2	12	32	32	3	2	rejected 👻				
datestamp: 2015-07-09	UK33		56	4	3	2	12	12	rejected 👻				
status: pending	UK34	76	3	5	3	87	43	34	rejected 👻				
									Update				
Messages									- Archive				
Timestamp User	M	essage	•						Archive of submission and any supporting files:				
2015-07-09 12:55:23+00 Keith Jolley You need to submit so				ates d	lata fo	r thes	e prot	iles.					
							-	_	Download TAR				
		Mess	age:	App	pend	s	end n	ow	Close submission				

### 7.3 Isolates

Clicking the appropriate submission on the 'Manage submissions' page.

Print/LST         Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications         Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Logged in: Keith Jolley (keith). @Log out   Change password	Help 🖉
Manage submissions	
Submit new data	
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.	
Submission type:	
isolates	
Pending submissions	
You have submitted the following submissions that are pending curation:	
Submission id         Submitted         Updated         Type         Details           BIGSdb_20150709121747_1342_99624         2015-07-09         2015-07-09         isolates         2	
Return to index page	

You will see a table summarizing the submission.

(PubMLST) Bre		rofile/ST   List Scheme/alleles   Publications ptions   Profiles/sequences d		ibase subi	missions						
Logged in: Keith Jolley (keit	h). ⊕Log out   Change passv	word									Help 🖓
Curate submis	ssion										
Submission: DIC 6d	b 00150700101747	4240 00004									
Submission: BIGSd	0_20150709121747	_1342_33624									
Summary											
type:	isolates										
submitter:	Joe Bloggs, University (	of Oxford, UK									
datestamp:	2015-07-09										
status:	pending										
Isolates											
isolate country		source species	serogroup					FetA_VR F1-5			
UK322 UK men UK325 UK	septicaemia	blood Neisseria meningitio		2 3 2 3	4 3 4 3		4 6 4 6	F1-5 F1-5	5 5-1	2	
Batch curate	oopiloaonna	oor nelooena meningia		2 0		10			nding 👻	Update	
Batch curate							Record	status, pe	nuing 👻	Opuale	
Messages	Arc	chive									
	Archiv	ive of submission and any sup	oporting files:								
	Dowr	nload 📊									
	.::										
Message: Append	Send now										

Click the 'Batch curate' button.

(PubMLST) Br	uery: Search   Browse   Pr reakdown: Isolate fields   3 nks: Contents   Home   O	Scheme/allele		nitions   Data	ibase s	ubmiss	ions						
Logged in: Keith Jolley (ke	ith). ເ⇔Log out   Change passv	vord											Help 🖉
Curate subm	ission												
Submission: BIGS	db 20150709121747	1342 9962	4										
	isolates	40-4											
	<ul> <li>Joe Bloggs, University</li> <li>2015-07-09</li> </ul>	or Oxford, UK											
	pending												
Isolates													
isolate country		source	species								R PorA_VR1		
UK322 UK me UK325 UK	ningitis and septicaemia septicaemia		seria meningitidis seria meningitidis	B	2	3 4	3	8	4	6 F1-5		2	
Batch curate			j						Red		pending 👻	Update	
Messages		hive	ion and any suppo	uting files:									
		_	on and any suppo	nung mes.									
	Dowr	nload TAR											
Message: Append	Send now												

This will take you to the batch isolate upload page in the curators' interface.

#### The upload form will be filled with details from the submission.

Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert							
Logged in: Keith Jolley (keith). @Log out   Change password	Toggle: 🕄						
Batch insert isolates							
This page allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.							
<ul> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.</li> <li>Enter references for your isolates as a semi-colon (;) separated list of PubMed ids (non-integer ids will be ignored).</li> <li>You can also upload allele fields along with the other isolate data - simply create a new column with the locus name (see the 'allowed_loci' tab in the Excel template for locus names). These will be added with a confirmed status and method set as 'manual'.</li> <li>You can choose whether or not to include an id number field - if it is omitted, the next available id will be used automatically.</li> <li>Download tab-delimited header for your spreadsheet - use 'Paste Special <sup>O</sup> Text' to paste the data.</li> </ul>							
Download submission template (xlsx format)							
Please select the sender from the list below:							
Bloggs, Joe (jbloggs) Value will be overridden if you include a sender field in your pasted data.							
Paste in tab-delimited text (include a field header line).	Action						
isolate country year aroE fumC gdh UK233 UK 2015 meningitis and septicaemia CSF Neisseria meningitidis B 2 3 4 3 8 4 6 F1-5 5 2 UK322 UK 2014 meningitis blood Neisseria meningitidis B 2 3 17 3 8 4 6 F1-5 5 2-1	Reset Submit						
Back							

Click submit to check and then import if there are no errors.

F	0
Database: Species home   Curator's page (species)   Curator's page (database) Users: Add   Query/update Isolates: Add   Query/update   Batch insert	
Logged in: Keith Jolley (keith). @Log out   Change password	Toggle: 🕄
Batch insert isolates	
Database updated ok Return to submission Back to main page	

After upload, a link on the confirmation page leads back to the submission management page.

**Note:** Depending on the database policy, definitions of new scheme profiles, e.g. for MLST, may require submission of representative isolate records. Where this is the case, the curator will need to extract the new profile from the submitted record. The tab-delimited isolate text file can be downloaded from the archive of supporting files linked to the submission and used directly for *batch adding new profiles*. Alternatively, the curator could use the *Export functionality* of the database to generate the file required for batch profile definition after upload of the isolate data.

#### 7.3.1 Requesting additonal information

You can send a message to the submitter by entering it in the Messages box and clicking 'Send now'. This will append a message to the submission and send an update to the submitter so that they can respond.

#### 7.3.2 Closing the submission

You can add a message to the submitter by entering it in the message box and clicking 'Append'. Change the record status to either 'accepted' or 'rejected' depending on whether you have accepted the submission. Click 'Update'.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Logged in: Keith Jolley (keith). (&Log out   Change password	Help 🖉
Curate submission	
Submission: BIGSdb_20150709121747_1342_99624	
type: isolates	
submitter: Joe Bloggs, University of Oxford, UK	
datestamp: 2015-07-09	
status: pending	
- Isolates	
isolate country  disease  source  species  serogroup abcZ adk aroE fumC gdh pdhC pgm FetA_VR PorA_VR1 PorA_VR2	
UK322 UK meningitis and septicaemia blood Neisseria meningitidis B 2 3 4 3 8 4 6 F1-5 5 2	
UK325 UK septicaemia CSF Neisseria meningitidis B 2 3 4 3 18 4 6 F1-5 5-1 2	
Batch curate Record status: accepted - Update	
Messages	
Archive of submission and any supporting files:	
Download TAR	
a bounded may	
Message: Append Send now	

The 'Close submission' button will now appear. Click this to close the submission. The submitter will be notified of their submission status.

(PubMLST) Breakdow		ST   List me/alleles   Publications s   Profiles/sequences defir	nitions   Data	abase s	submis	sions							
Logged in: Keith Jolley (keith). 🕪	og out   Change password												Help 🖉
Curate submissi	on												
Submission: BIGSdb 20	0150709121747 134	2 99624											
Summary													
type: isola	ates Bloggs, University of Ox	and LIK											
datestamp: 2015		ord, OK											
status: pend													
-Isolates													
	disease sour	ce species od Neisseria meningitidis		abcZ 2		0E   fum() 4 3	gdh 8	pdhC 4		A_VR 1-5	PorA_VR1 5	PorA_VR2 2	
	pticaemia CS			2	-	4 3	18	4		1-5	5-1	2	
								Re	cord statu	is: ac	cepted 👻	Update	
Messages	Archive			— Act	tion —								
	Archive of	submission and any suppo	orting files:	_		mission							
	Download			0.0									
	.::												
Message: Append Se	nd now												

### **Offline curation tools**

### 8.1 Automated offline sequence tagging

Sequence tagging is the process of identifying alleles by scanning the sequence bin linked to an isolate record. Loci need to be defined in an external sequence definition database that contains the sequences for known alleles. The tagging function uses BLAST to identify sequences and will tag the specific sequence region with locus information and an allele designation if a matching allele is identified by reference to an external database.

There is a script called 'autotag.pl' in the BIGSdb package. This can be used to tag genome sequences from the command line.

Before autotag.pl can be run for the first time, a log file needs to be created. This can be created if it doesn't already exist with the following:

```
sudo touch /var/log/bigsdb_scripts.log
sudo chown bigsdb /var/log/bigsdb_scripts.log
```

The autotag.pl script should be installed in /usr/local/bin. It is run as follows:

autotag.pl --database <database configuration>

where <database configuration> is the name used for the argument 'db' when using the BIGSdb application.

If you have multiple processor cores available, use the –threads option to set the number of jobs to run in parallel. Isolates for scanning will be split among the threads.

The script must be run by a user that can both write to the log file and access the databases, e.g. the 'bigsdb' user (see 'Setting up the offline job manager').

A full list of options can be found by typing:

```
autotag.pl --help
NAME
   autotag.pl - BIGSdb automated allele tagger
SYNOPSIS
   autotag.pl --database NAME [options]
OPTIONS
-0, --missing
   Marks missing loci as provisional allele 0. Sets default word size to 15.
-d, --database NAME
```

```
Database configuration name.
-h, --help
   This help page.
-i, --isolates LIST
   Comma-separated list of isolate ids to scan (ignored if -p used).
--isolate_list_file FILE
  File containing list of isolate ids (ignored if -i or -p used).
-I, --exclude_isolates LIST
   Comma-separated list of isolate ids to ignore.
-1, --loci LIST
   Comma-separated list of loci to scan (ignored if -s used).
-L, --exclude_loci LIST
    Comma-separated list of loci to exclude
-m, --min_size SIZE
   Minimum size of seqbin (bp) - limit search to isolates with at least this
   much sequence.
-n, --new_only
   New (previously untagged) isolates only. Combine with --new_max_alleles
    if required.
--new_max_alleles ALLELES
   Set the maximum number of alleles that can be designated or sequences
   tagged before an isolate is not considered new when using the --new_only
   option.
-o, --order
   Order so that isolates last tagged the longest time ago get scanned first
    (ignored if -r used).
--only_already_tagged
   Only check loci that already have a tag present (but no allele
   designation). This must be combined with the --already_tagged option or
   no loci will match. This option is used to perform a catch-up scan where
   a curator has previously tagged sequence regions prior to alleles being
   defined, without the need to scan all missing loci.
-p, --projects LIST
    Comma-separated list of project isolates to scan.
-P, --exclude_projects LIST
   Comma-separated list of projects whose isolates will be excluded.
-q, --quiet
   Only error messages displayed.
-r, --random
   Shuffle order of isolate ids to scan.
-R, --locus_regex REGEX
   Regex for locus names.
```

```
-s, --schemes LIST
   Comma-separated list of scheme loci to scan.
-t, --time MINS
    Stop after t minutes.
--threads THREADS
   Maximum number of threads to use.
-T, --already_tagged
    Scan even when sequence tagged (no designation).
-v, --view VIEW
   Isolate database view (overrides value set in config.xml).
-w, --word_size SIZE
   BLASTN word size.
-x, --min ID
   Minimum isolate id.
-y, --max ID
   Maximum isolate id.
```

### 8.2 Automated offline allele definition

There is a script called 'scannew.pl' in the BIGSdb scripts/automation directory. This can be used to identify new alleles from the command line. This can (optionally) upload these to a sequence definition database.

Before scannew.pl can be run for the first time, a log file needs to be created. This can be created if it doesn't already exist with the following:

```
sudo touch /var/log/bigsdb_scripts.log
sudo chown bigsdb /var/log/bigsdb_scripts.log
```

The autotag.pl script should be installed in /usr/local/bin. It is run as follows:

scannew.pl --database <database configuration>

where <database configuration> is the name used for the argument 'db' when using the BIGSdb application.

If you have multiple processor cores available, use the –threads option to set the number of jobs to run in parallel. Loci for scanning will be split among the threads.

The script must be run by a user that can both write to the log file and access the databases, e.g. the 'bigsdb' user (see 'Setting up the offline job manager').

A full list of options can be found by typing:

```
scannew.pl --help
NAME
scannew.pl - BIGSdb automated allele definer
SYNOPSIS
scannew.pl --database NAME [options]
```

OPTIONS -a, --assign Assign new alleles in definitions database. --allow\_frameshift Allow sequences to contain a frameshift so that the length is not a multiple of 3, or an internal stop codon. To be used with --coding\_sequences option to allow automated curation of pseudogenes. New alleles assigned will be flagged either 'frameshift' or 'internal stop codon' if appropriate. Essentially, combining these two options only checks that the sequence starts with a start codon and ends with a stop codon. -A, --alignment INT Percentage alignment (default: 100). -B, --identity INT Percentage identity (default: 99). -c, --coding\_sequences Only return complete coding sequences. -d, --database NAME Database configuration name. -h, --help This help page. -i, --isolates LIST Comma-separated list of isolate ids to scan (ignored if -p used). --isolate\_list\_file FILE File containing list of isolate ids (ignored if -i or -p used). -I, --exclude\_isolates LIST Comma-separated list of isolate ids to ignore. -l, --loci LIST Comma-separated list of loci to scan (ignored if -s used). -L, --exclude loci LIST Comma-separated list of loci to exclude. -m, --min\_size SIZE Minimum size of seqbin (bp) - limit search to isolates with at least this much sequence. -n, --new\_only New (previously untagged) isolates only. -o, --order Order so that isolates last tagged the longest time ago get scanned first (ignored if -r used). -p, --projects LIST Comma-separated list of project isolates to scan. -P, --exclude\_projects LIST

```
Comma-separated list of projects whose isolates will be excluded.
-r, --random
    Shuffle order of isolate ids to scan.
-R, --locus_regex REGEX
   Regex for locus names.
-s, --schemes LIST
   Comma-separated list of scheme loci to scan.
-t, --time MINS
   Stop after t minutes.
--threads THREADS
   Maximum number of threads to use.
-T, --already_tagged
    Scan even when sequence tagged (no designation).
-v, --view VIEW
    Isolate database view (overrides value set in config.xml).
-w, --word_size SIZE
   BLASTN word size.
-x, --min ID
   Minimum isolate id.
   --max ID
-v,
   Maximum isolate id.
```

### 8.3 Cleanly interrupting offline curation

Sometimes you may wish to stop running autotagger or allele autodefiner jobs as they can be run for a long time and as CRON jobs. If these are running in single threaded mode, the easiest way is to simply send a kill signal to the process, i.e. identify the process id using 'top', e.g. 23232 and then

kill 23232

The scripts should respond to this signal within a couple of seconds, clean up all their temporary files and write the history log (where appropriate). Do not use 'kill -9' as this will terminate the processes immediately and not allow them to clean up.

If these scripts are running using multiple threads, then you need to cleanly kill each of these. The simplest way to terminate all autotagger jobs is to, type

pkill autotag

The parent process will wait for all forked processes to cleanly terminate and then exit itself.

Similarly, to terminate all allele autodefiner jobs, type

pkill scannew

### 8.4 Uploading contigs from the command line

There is a script called upload\_contigs.pl in the BIGSdb scripts/maintenance directory. This can be used to upload contigs from a local FASTA file for a specified isolate record.

The upload\_contigs.pl script should be installed in /usr/local/bin. It is run as follows:

```
upload_contigs.pl --database <NAME> --isolate <ID> --file <FILE>
--curator <ID> --sender <ID>
```

The script must be run by a user who has the appropriate database permissions and the local configuration settings should be modified to match the database user account to be used. The default setting uses the 'apache' user which is used by the BIGSdb web interface.

A full list of options can be found by typing:

```
upload_contigs.pl --help
NAME
    upload_contigs.pl - Upload contigs to BIGSdb isolate database
SYNOPSIS
    upload_contigs.pl --database NAME --isolate ID --file FILE
         --curator ID --sender ID [options]
OPTIONS
-a, --append
   Upload contigs even if isolate already has sequences in the bin.
-c, --curator ID
   Curator id number.
-d, --database NAME
   Database configuration name.
-f, --file FILE
    Full path and filename of contig file.
-h, --help
   This help page.
-i, --isolate ID
    Isolate id of record to upload to.
-m, --method METHOD
   Method, e.g. 'Illumina', default 'unknown'.
--min_length LENGTH
   Exclude contigs with length less than value.
-s, --sender ID
    Sender id number.
```

### **Definition downloads**

The sequence definition database defines alleles, i.e. links an allele identifier to a sequence. It also defines scheme, e.g. MLST, profiles.

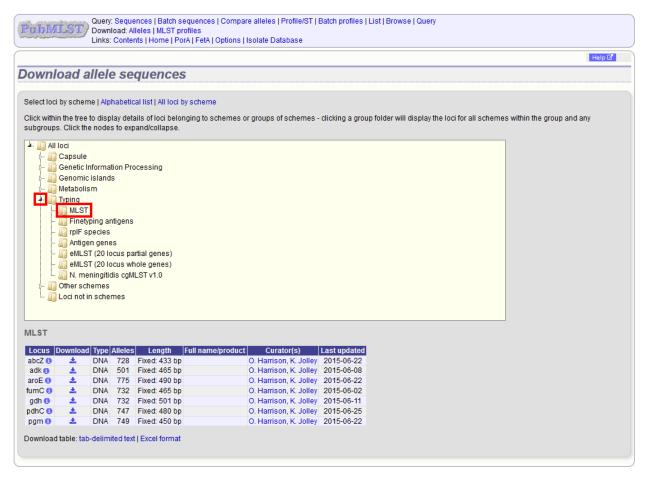
### 9.1 Allele sequence definitions

Click the 'Allele sequences' link in the 'Downloads' section. Depending on the database, you may see either a hierarchical scheme tree or a table of loci. You can choose to display links either by scheme using the scheme tree, as an alphabetical list or a page of all schemes, by selecting the appropriate link at the top of the page.

#### 9.1.1 Scheme tree



You can drill down through the tree by clicking branch nodes. Clicking the labels of internal nodes will display tables of all schemes belonging to that scheme group. Clicking the labels of terminal nodes will display that single scheme table.



Click the download link for the required locus

Locus	Download	Туре	Alleles	Length	Full name/product	Curator(s)	Last updated
abcZ 🚺	*	DNA	728	Fixed: 433 bp		O. Harrison, K. Jolley	2015-06-22
adk 📵	<u>*</u>	DNA	501	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-08
aroE 🕕	*	DNA	775	Fixed: 490 bp		O. Harrison, K. Jolley	2015-06-22
fumC 📵	*	DNA	732	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-02
gdh 🚺	*	DNA	732	Fixed: 501 bp		O. Harrison, K. Jolley	2015-06-11
pdhC 🕕	*	DNA	747	Fixed: 480 bp		O. Harrison, K. Jolley	2015-06-25
pgm 🕕	٠.	DNA	749	Fixed: 450 bp		O. Harrison, K. Jolley	2015-06-22

#### Alleles will be downloaded in FASTA format, e.g.

#### >fumC\_1

GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCGTCGGCATTGAACCCGTACCGGAAAAAATCGACTATTTCCTGCACCATTCC
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAC
>fumC_2
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGG
TTGGGCGAAATCAAAATCCCCGAAAACGAGCCGGGTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGCATGGCGGGCGCGCCGGGCAATTTCGAGCTGAACGTCTATATGCCCGTTATC
GCCTACAACCTCTTGCAATCCATCCGCCTCTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCATCGGCATCGAACCCGTACCGGAAAAAATCGACTATTTCCTGCACCATTCC
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAC
>fumC_3
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGG
TTGGGCGAAATCAAAATCCCCGAAAACGAGCCGGGTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGCATGGCGGGCGCGCCGGGCAATTTCGAGCTGAACGTCTATATGCCCGTTATC
GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCGTCGGCATCGAACCCGTACCGGAAAAAATCGACTATTTCCTGCACCATTCC
CTGATGCTGGTTACTGCGTTAAAACCGTAAAAATCGGCTACGAAAAAC

### 9.1.2 Alphabetical list

Loci can be displayed in an alphabetical list. Loci will be grouped in to tables by initial letter. If common names are set for loci, they will be listed by both primary and common names.

Li	nks: Cont	ents   H	lome   I	PorA   FetA   Options   Isolate Dat	tabase		
							He
ownload all	ele se	eaue	ence	25			
elect loci by scheme	Alphabol	lical list		i by achoma			
election by scheme	Alphabel	ucarnst	PAILIO	a by scheme			
Locus	Download		Alleles	Length	Full name/product	Curator(s)	Last updated
ArsR [NEIS1769] ()	*	DNA	70	Variable: No limits set	ArsR family transcriptional regulator		2015-03-18
AsnC [NEIS1566] ()	±	DNA	160	Variable: No limits set	transcription regulator AsnC		2015-05-20
abcZ 🕕	±	DNA	728	Fixed: 433 bp		O. Harrison, K. Jolley	2015-06-22
abcZ [NEIS1015] 🕕	±	DNA	493	Variable: No limits set	ABC transporter ATP-binding protein		2015-05-19
aceF [NEIS1279] ()	٤.	DNA	477	Variable: (1563 min; 1641 max)	dihydrolipoamide acetyltransferase (EC 2.3.1.12)		2015-05-19
ackA2 [NEIS1727] 🕕	±	DNA	408	Variable: No limits set	acetate kinase		2015-05-20
acnA [NEIS1729] 🕕	±	DNA	527	Variable: No limits set	aconitate hydratase 1 (EC 4.2.1.3)		2015-05-20
acnB [NEIS1492] ()	±	DNA	476	Variable: No limits set	aconitate hydratase 2 (EC 4.2.1.3)		2015-05-19
adhA [NEIS0486] ()	±	DNA	614	Variable: No limits set	alcohol dehydrogenase		2015-05-17
adhC [NEIS1241] 🕦	±	DNA	239	Variable: No limits set	alcohol dehydrogenase		2015-05-20
adk 🕕	±	DNA	501	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-08
adk [NEIS0767] ()	*	DNA	208	Variable: No limits set	adenylate kinase		2015-05-18
aldA [NEIS1942] 🕕	*	DNA	447	Variable: No limits set	aldehyde dehydrogenase		2015-05-20
aniA [NEIS1549] 🕕		DNA	0	Variable: No limits set	nitrite reductase, major outer membrane copper-containing protein		
anmK [NEIS1788] 🕕	*	DNA	480	Variable: No limits set	anhydro-N-acetylmuramic acid kinase	A. Jamet	2015-05-20
apaH [NEIS0610] 🕦	±	DNA	116	Variable: No limits set	diadenosine tetraphosphatase		2015-05-17
argH [NEIS0580] 🕕	*	DNA	433	Variable: No limits set	argininosuccinate lyase		2015-05-18
aroE 🕕	*	DNA	775	Fixed: 490 bp		O. Harrison, K. Jolley	2015-06-22
aroE [NEIS1810] 🕕	*	DNA	305	Variable: No limits set	shikimate dehydrogenase		2015-05-21
aspA 🕕	±	DNA	189	Fixed: 432 bp		K. Jolley	2015-04-21
aspA [NEIS1185] 🕕	±	DNA	501	Variable: No limits set	aspartate ammonia-lyase		2015-05-20
atIA [NEIS2274] ()	<u>*</u>	DNA	18	Variable: No limits set	atlA / peptidoglycan transglycosylase	O. Harrison	2014-12-10
autA [NEIS1859] 1	*	DNA	242	Variable: No limits set	autotransporter A		2015-05-20

Click the download links for the required locus.

#### 9.1.3 All loci by scheme

Loci can also be displayed by scheme with all schemes displayed.

PubM	LST	Downl	oad: Alle	eles   ML	atch sequences   Cor LST profiles ne   PorA   FetA   Optio			profiles   List   Browse   Query			
											Help 🗹
Downlo	oad a	llele	e seg	uen	ices						
Select loci I	by schem	ne   Alph	nabetica	Il list   Al	Il loci by scheme						
MLST											
IVILOI											
Locus D				Leng				updated			
abcZ 🕕	*	DNA		Fixed: 43		O. Harrison, K.					
adk 🕕	*	DNA		Fixed: 4		O. Harrison, K.					
aroE 🕕	*			Fixed: 4		O. Harrison, K.					
fumC 🚯	*			Fixed: 4	•	O. Harrison, K.	-				
gdh 🕕	*	DNA		Fixed: 5	•	O. Harrison, K.					
pdhC 🕕	*	DNA		Fixed: 4	•	O. Harrison, K.					
pgm 🚺	*	DNA	749 1	Fixed: 4	50 DD	O. Harrison, K.	Jolley 201	5-06-22			
Finetypin	na antio	aens									
Locus	Down			lleles	Length	Full name/produc		(s) Last updated			
PorA VR1					/ariable: No limits set	-		*			
PorA VR2	-				/ariable: No limits set						
FetA VR 👩	) ±	, pe	ptide	581 V	/ariable: No limits set		I. Feave	rs 2015-06-26			
rpIF spec	ries										
ipii spec	0100										
Locus Do				Lengt				ast updated			
'rpIF 📵	±	DNA	123 Fi	ixed: 41	3 bp 50S ribosomal p	protein L6 (partial) E	Watkins	2015-01-30			
ADP-hep	tose bi	osynti	nesis								
Loc	cus	Dow	nload T	vpe Alle	eles Length		Full na	me/product	Aliases	Curator(s)	Last updated
NEIS0769		_			44 Variable: No limi	ts set D-be		e-7-phosphate kinase	NMB0825; NMC0769; rfaE		2015-05-18
NEIS0773	B (hldD)	)	± D	DNA 3	11 Variable: No limi			heptose epimerase	NMB0828; NMC0773		2015-05-17
NEIS2014				DNA 2	16 Variable: No limi			1,7 bisphosphate phosphatase		C. Kahler	
NEIS2055					06 Variable: No limi			osphate adenvivitransferase	NMB2076; NMC2055		2015-05-21

Click the green download links for the required locus.

#### 9.1.4 Download locus table

The locus table can be downloaded in tab-delimited text or Excel formats by clicking the links following table display.

	Genetic	Informati	ion Pro	ocessing				
ķ. 🌆	Genomi	c islands	3					
- k- 🛄 I	Metabol	ism						
- 👍 🛄 '	Typing							
i F,	🔲 MLS	т						
	🔲 Finet	typing an	tigens					
	🔲 rpIF :	species						
	Antig	en gene	s					
	eML	ST (20 lo	cus pa	rtial genes)				
				nole genes)				
				LST v1.0				
	Other so	-	lo ogin	201110				
		in schen	nac					
	Locinot	In schen	lica					
MLST								
Locus D	)ownioa	d Type /	Alleles	Length	Full name/product Curator(s)	Last updated		
Locus D abcZ 1	)ownloa 土	d Type / DNA		Length Fixed: 433 bp		Last updated 2015-06-22		
		DNA	728		O. Harrison, K. Jolley	2015-06-22		
abcZ 🕕	±	DNA	728 501	Fixed: 433 bp	O. Harrison, K. Jolley O. Harrison, K. Jolley	2015-06-22 2015-06-08		
abcZ () adk ()	* *	DNA DNA DNA	728 501	Fixed: 433 bp Fixed: 465 bp	O. Harrison, K. Jolley O. Harrison, K. Jolley O. Harrison, K. Jolley O. Harrison, K. Jolley	2015-06-22 2015-06-08 2015-06-22		
abcZ () adk () aroE () fumC () gdh ()	* * * * *	DNA DNA DNA DNA DNA	728 501 775 732 732	Fixed: 433 bp Fixed: 465 bp Fixed: 490 bp	O. Harrison, K. Jolley O. Harrison, K. Jolley O. Harrison, K. Jolley O. Harrison, K. Jolley O. Harrison, K. Jolley	2015-06-22 2015-06-08 2015-06-22 2015-06-02 2015-06-11		
abcZ () adk () aroE () fumC ()	* * *	DNA DNA DNA DNA DNA DNA	728 501 775 732 732 747	Fixed: 433 bp Fixed: 465 bp Fixed: 490 bp Fixed: 465 bp Fixed: 465 bp Fixed: 501 bp Fixed: 480 bp	O. Harrison, K. Jolley O. Harrison, K. Jolley	2015-06-22 2015-06-08 2015-06-22 2015-06-02 2015-06-11		
abcZ () adk () aroE () fumC () gdh ()	* * * * *	DNA DNA DNA DNA DNA DNA	728 501 775 732 732 747	Fixed: 433 bp Fixed: 465 bp Fixed: 490 bp Fixed: 465 bp Fixed: 501 bp	O. Harrison, K. Jolley O. Harrison, K. Jolley	2015-06-22 2015-06-08 2015-06-22 2015-06-02 2015-06-11 2015-06-25		
abcZ () adk () aroE () fumC () gdh () pdhC () pgm ()	* * * * * * *	DNA DNA DNA DNA DNA DNA DNA	728 501 775 732 732 747 749	Fixed: 433 bp Fixed: 465 bp Fixed: 490 bp Fixed: 465 bp Fixed: 465 bp Fixed: 501 bp Fixed: 480 bp Fixed: 450 bp	O. Harrison, K. Jolley O. Harrison, K. Jolley	2015-06-22 2015-06-08 2015-06-22 2015-06-02 2015-06-11 2015-06-25		
abcZ () adk () aroE () fumC () gdh () pdhC () pgm ()	* * * * * * *	DNA DNA DNA DNA DNA DNA DNA	728 501 775 732 732 747 749	Fixed: 433 bp Fixed: 465 bp Fixed: 490 bp Fixed: 465 bp Fixed: 465 bp Fixed: 501 bp Fixed: 480 bp	O. Harrison, K. Jolley O. Harrison, K. Jolley	2015-06-22 2015-06-08 2015-06-22 2015-06-02 2015-06-11 2015-06-25		

# 9.2 Scheme profile definitions

Scheme profiles, e.g. those for MLST, can be downloaded by clicking the appropriate link on the contents page.

Prrfn/ILST         Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Bro           Download: Alleles   MLST profiles         Links: Contents   Home   Options   PubMLST.org   Isolate Database	owse   Query
Campylobacter locus/sequence definitions database The Campylobacter PubMLST sequence definition database contains allele and profile data representing the total database should have a corresponding record in the isolate database.	I known diversity of <i>C. jejuni</i> and <i>C. coli.</i> Every new ST deposited in this
Query database <ul> <li>Sequence query - query an allele sequence.</li> <li>Batch sequence query - query multiple sequences in FASTA format.</li> <li>Sequence attribute search - find alleles by matching attributes.</li> <li>Browse MLST profiles</li> <li>Search MLST profiles matched to entered list.</li> <li>Search by combinations of MLST alleles - including partial matching.</li> <li>Batch profile query - lookup MLST profiles copied from a spreadsheet.</li> </ul>	<ul> <li>Option settings</li> <li>Set general options</li> <li>Submissions</li> <li>Manage submissions</li> <li>General information</li> <li>Number of sequences: 412175</li> <li>Number of profiles (MLST): 7841</li> <li>Last updated: 2015-06-17</li> <li>Profile update history</li> <li>About BIGSdb</li> </ul>
Export     Sequences - XMFA / concatenated FASTA formats     Sequence similarity - find sequences mos     Sequence comparison - display a comparison - disp	rison between two sequences.

If multiple schemes are available, you will need to select the scheme in the dropdown box and click 'Download profiles'

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database								
Toggle: ()								
Neisseria locus/sequence definitions database								
The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.								
<ul> <li>Query database</li> <li>Sequence query - query an allele sequence.</li> <li>Batch sequence query - query multiple sequences in FASTA format.</li> <li>Batch sequence query - query multiple sequences in FASTA format.</li> <li>Browse profiles</li> <li>Search profiles</li> <li>Search profiles matched to entered list.</li> <li>Search profiles copied from a spreadsheat.</li> <li>Extract finetype from whole genome data</li> </ul> Option settings <ul> <li>Allele sequences: 519864</li> <li>Number of profiles: Show</li> <li>Last updated: 2015-06-26</li> <li>Profile update history</li> <li>About BICS3db</li> </ul>								
Export • Sequences - XMFA / concatenated FASTA formats Analysis • Sequence similarly - find sequences most similar to selected allele. • Sequence comparison - display a comparison between two sequences. • Locus Explorer - tool for analysing allele sequences stored for particular locus.								

Profiles will be downloaded in tab-delimited format, e.g.

ST	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	clonal_complex
1	1	3	1	1	1	1	3	ST-1 complex/subgroup I/II

2	1	3	4	7	1	1	3	ST-1 complex/subgroup I/II
3	1	3	1	1	1	23	13	ST-1 complex/subgroup I/II
4	1	3	3	1	4	2	3	ST-4 complex/subgroup IV
5	1	1	2	1	3	2	3	ST-5 complex/subgroup III
6	1	1	2	1	3	2	11	ST-5 complex/subgroup III
7	1	1	2	1	3	2	19	ST-5 complex/subgroup III
8	2	3	7	2	8	5	2	ST-8 complex/Cluster A4
9	2	3	8	10	8	5	2	ST-8 complex/Cluster A4
10	2	3	4	2	8	15	2	ST-8 complex/Cluster A4
11	2	3	4	3	8	4	6	ST-11 complex/ET-37 complex
12	4	3	2	16	8	11	20	
13	4	10	15	7	8	11	1	ST-269 complex
14	4	1	15	7	8	11	1	ST-269 complex

### **Data records**

Record pages for different types of data can be accessed following a query by clicking appropriate hyperlinks.

### 10.1 Isolate records

An Isolate record page displays everything known about an isolate.

Full Information on Isolate M10 240474         Priorts         The Main process of the following projects:         MEM Main process of Genome Library is a cell loading the bareen Public Health England, The Wellcome Trust Barger Institute and the University of Oxford, Uxr (Importantion)         O MME         Memory User Research Fournation:         D MME Memory Colspan="2">Memory is a cell loading to optimate the last the England, The Wellcome Trust Barger Institute and the University of Oxford, Uxr (Importantion)         D MME Memory Colspan="2">Memory Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2">Colspan="2"         MET-MOL loadies up to part 20102011 to 2011/2012 excluding Northern Instand         MET-MOL loadies up to 20102011 to 2011/2012 excluding Northern Instand         MET-MOL loadies up to 2010201 to 2011/2012 excluding Northern Instand         MET-MOL loadies up to 2010201 to 2011/2012 excluding Northern Instand         MET-MOL loadies up to 2010201         Colspan="2"         MET-MOL loadies up to 2010201 to 2011/2012 excluding Northern Instand         Colspan="2"         Colspan="2"         Colspan= 20102011         Colspan="2"         Colspan= 2010201<	Ouery: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions								
The isolate is a member of the following projects: INFR Manifogacoccas Genome Library is a caliboration between Public Health England, The Wallcome Trast Surger Institute and the University of Automation and Automatica and A	Full informat	ion on iso	late M10 240474						
NEF Maingococcus Genome Library is a collaboration between Public Health England. The Welfcome Truat Surger Institute and the University of Oxford, Mind by the Memory Beacement Foundation.         Description of the Wenger Steerage Provide State Steerage Steerage State Steerage Steerage State Steerage Steerage Steerage State Steerage Ste	Projects								
The MFF Meningsoccas Genome Library is a caliboration between Public Health England, The Weldsome Trust Sanger Institute and the University of Set of the MFF Genome Library data mult be cited in any publication or presentation making use of it.	This isolate is a member of the following projects:								
kotkd, funded by her Meninglis Research Foundation.       Second Foundation.         Second Foundation.       Second Foundation.         MRF-MGL Isolates epi years. 2010;2011 to 2011/2012 excluding Northern Ireland       EAW_genography. J.M.RF         All NRFF genomes excluding Northern Ireland       MRF_no_N         All NRFF genomes excluding Northern Ireland       MRF_no_N         Provenance/meta data       seconder.         dis 1986       species: Networks mail         isolate:       M10/24074       seconder.         contings:       2010       ENA accession:         generation:       beroket.       Seconder.         generation:       Seconder.       Seconder.									
				ngland, The Wellcome Trust Sar	ger Institute and the University of	A			
B29_MRF         MRF-MGL isolates eqi years 2010/2011 to 2011/2012 excluding Northem Ireland         EAW_genogroup B_MRF         All MRF Meningococcus Genome Library genogroup B isolates from England and Wales in 2010/11 and 2011/12         MRF_mo_NI         All MRF Meningococcus Genome Library genogroup B isolates from England and Wales in 2010/11 and 2011/12         MRF_mo_NI         All MRF genomes excluding Northem Ireland         Provenance/meta data         isolates M10/20074         Sequence Bil         country: UK	Use of the MRF Genon	ne Library data must	be cited in any publication or presentation	making use of it.		Meningitis			
EAW_genogroup_B_MRF         All MRF. Meningcoccus Genome Library genogroup B isolates from England and Wales in 2010/11 and 2011/12         MRE_no_NI         All MRF genomes excluding Northem Ireland         Colspan="2">MRE_no_NI         Colspan= MRE_N Colspan="2">Colspan= MRE_N Colspan= MRE_N Colspan="2">Colspan= MRE_N Colspan= MRE						Research Foundation			
EAW_genorpoup_B_MRF         IN MRF_no_NI         MRF_no_NI         All MRF genomes excluding Northern Ireland         MRF_no_NI         All MRF genomes excluding Northern Ireland         MRF_no_NI         All MRF genomes excluding Northern Ireland         Provenance/meta data         idit 1998         idit 1998         Strain designation: B: P1:15:11; F3:9: ST-269 (cc289)         ENA accession: ERR/060224 moves datase         country: UK         country: UK         country: UK (E-mail: dordred hill)         country: UK (E-mail: dordred hill) <td col<="" td=""><td></td><td></td><td></td><td>899_MRF</td><td></td><td></td></td>	<td></td> <td></td> <td></td> <td>899_MRF</td> <td></td> <td></td>				899_MRF				
All MRF Metringococcus Genome Library genogroup B isolates from England and Wales in 2010/11 and 2011/12           MRF_no_NI           All MRF genomes excluding Northern Ireland           MRF_no_NI           MRE           Colspan="2"           MRF_no_NI           MRF_no_NI           MRE <t< td=""><td>MRF-MGL isolates epi</td><td>years 2010/2011 to</td><td>2011/2012 excluding Northern Ireland</td><td></td><td></td><td></td></t<>	MRF-MGL isolates epi	years 2010/2011 to	2011/2012 excluding Northern Ireland						
All MRF Metringococcus Genome Library genogroup B isolates from England and Wales in 2010/11 and 2011/12           MRF_no_NI           All MRF genomes excluding Northern Ireland           MRF_no_NI           MRE           Colspan="2"           MRF_no_NI           MRF_no_NI           MRE <t< td=""><td></td><td>_</td><td>E&amp;W</td><td>_genogroup_B_MRF</td><td>_</td><td></td></t<>		_	E&W	_genogroup_B_MRF	_				
All MRF genomes excluding Northern Ireland Provenance/meta data id: 19988 species: Neisseria meningitidis isolat: M10 240474 serogroup: 8 strain designation: B: P119:1,15:11: F3-9: ST-299 (cc269) EINA accession: ERR06/224 www.destextal country: UK serogroup: 8 country: UK serogroup: 0 country: UK serogroup: 0 downleaded set country: UK serogroup: 0 country: UK serogroup: 0 downleaded set year: 2010 update: bloch light on accessing regidemiological year: 07/2010-06/2011 date entered: 2012-02-15 disease: invasive (unspecified/dher) date strain: 2014-06-23 Sequence bin contigs: 275 N90: 6405 total length: 109893 bp loci tagged: 1611 mean length: 7802 bp detailed breakdown: Display: N50: 34308 Sectemes and loci fund Copsule fund Copsu	All MRF Meningococci	us Genome Library (							
All MRF genomes excluding Northern Ireland Provenance/meta data id: 18968 species: Neisseria meningitidis isolate: M10 240474 serogroup: 8 strain designation: B: PI.19.1,511: F3-9: ST-299 (cc269) EINA accession: EFR060224 www.dca.su.st country: UK serogroup: 8 country: UK serogroup: 0 country: UK serogroup: 0 contract: 2012-01 disease: invasive (unspecified/other) data thistory: 100 spaces show details epidemiological year: 07/2010-06/2011 data entered: 2012-02-15 disease: invasive (unspecified/other) datastamp: 2014-06-23 Sequence bin contigs: 275 N99: 4405 total length: 103893 bp loci tagged: 1611 mean length: 7892 bp detailed breakdown: Display: N30: 3408 Schemes and loci wight biotic loci Information Processing Comparison Processing Pitting Consule Pitting Consule P	_								
Provenance/meta data         id:       19968       species:       Neisseria meningitidis         isolate:       M10 240474       serogroup:       B         strain designation:       B: P1.19.1,15.11; F3-9; ST-269 (cc269)       ENA accession:       ER060224       owned stats         continet:       Europe       curator:       Dorothea Hill       Ocorthea Hill       Ocorthea Hill         continet:       Europe       curator:       Dorothea Hill       Ocorthea Hill<		_		MRF_no_NI					
id:       1988       species:       Nelseria meningitidis         isolate:       M10 240474       serogroup:       B         strain designatio:       B: P1.19.1.15.11: F3.9: ST-269 (oc269)       ENA accessio:       ERR08624 arwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww	All MRF genomes excl	uding Northern Irela	nd						
id:       1988       species:       Nelseria meningitidis         isolate:       M10 240474       serogroup:       B         strain designation:       B: P1.19-1.15.11: F3-9: ST-269 (oc269)       ENA accession:       ERR086224 arwwww.strain         country:       UK       sender:       Dorothea Hill         country:       UK       sender:       Dorothea Hill         country:       South East       curator:       Dorothea Hill         continent:       Europe       curator:       Dorothea Hill       Outrot, UK (E-mail:         region:       South East       curator:       Dorothea Hill       Outrot, UK (E-mail:         region:       South East       update history:       100 goding show details         epidemiological yea:       072010-06/2011       date entered:       2012-02-15         disease:       invasive (unspecified/other)       date entered:       2012-02-15         disease:       1072010-06/2011       date entered:       2012-02-15         disease:       108059 bp       locat tagged:       1611         mean length:       1982 bp       detailed breakdown:       Display         No:       34308       seroserosina       Seroserosina         Secones and loci       Geneti									
isolate: M10240474 sergroup: B strain designation: B:P1.19-1,15-11: F3-9: ST-269 (cc269) ENA accession: ERR066224 _ xvvvvc5i fc.13 country: UK sender: Dorothea Hill continet: Europe country: UK continet: Europe 2010 update history: 100 weaking show details epidemiological year: 07/2010-06/2011 date entered: 2012-02-15 disease: invasive (unspecified/other) datestamp: 2014-06-23 Sequence bin contigs: 275 N90: 6405 total length: 2195045 5p N93: 3513 max length: 109859 5p doci tagged: 1611 mean length: 7982 5p detailed breakdown: Display: N50: 34308 Schemes and loci ↓ All loci ↓ All loci ↓ Matabolism: Josepsing Matabolism: Josepsing Matabolis	Provenance/meta	data							
strain designation       B: P1.19-1,15-11: F3-9: ST-269 (occ69)       ENA accession       ERR086224 → www.stated         country:       UK       sende:       Dorothea Hill         contine:       Europe       curate:       Dorothea Hill, University of Oxford, UK (E-mail:         region:       South East       curate:       Dorothea Hill, University of Oxford, UK (E-mail:         year:       2010       update history:       103 update; show details         epidemiological year:       07/2010-06/2011       date entered:       2012-02-15         disease:       invasive (unspecified/other)       date states:       2014-06-23         Sequence bin       contigs:       275       N90:       6405         total length:       2195045 bp       N90:       6405         total length:       1982 bp       loci tagged:       1611         mean length:       1982 bp       detailed breakdown:       Display         N50:       34308       Display       Statest       Statest	id:	18968		species:	Neisseria meningitidis				
contineIKsenderDorothea HillcontineEuropecurateDorothea Hill, University of Oxford, UK (E-nail: contine)regionSouth Eastupdate history103 updates, show detailsepidemiological yee07/2010-06/2011date entere2012-2-15diseaseinvasive (unspecified/other)date states2014-06-23Sequence bin2155535513max length19859 bploci tagged1611mean length752 bpdetalled breakdown:DisplayNSi3430834308Display									
continet:Europecurator:Dorothea Hill, University of Oxford, UK (E-mail: dorothea. hill@200.0X.ac.uk)region:South Eastcorthea. hill@200.0X.ac.uk)year:2010update histor:100 update show detailsepidemiological year:072010-06/2011date entered:2012-02-15disease:invasive (unspecified/other)date stamp:2014-06-23Sequence bin275N90:6405total length:195045 bpN90:6405invasive (unspecified/other)6405100state and length:195045 bpLoci taged:1611mean length:1982 bpdetailed breakdown:Displayx50:24308AutoriaDisplay:state and below:Sequence binSequence bininvasive display:Navigate and select schemes within three to display allele designationsSequence bininvasive display:Navigate and select schemes within three to display allele designationsSequence bininvasive display:Navigate and select schemes within three to display allele designationsSequence bininvasive display:Navigate and select schemes within three to display allele designationsSequence bininvasive display:Navigate and select schemes within three to display allele designationsSequence bininvasive display:Navigate and select schemes within three to display allele designationsSequence bininvasive display:Navigate and select schemes within three to display allele designationsSequence bininvasive display:<			-3-9: ST-269 (cc269)						
regin       South East       update history:       (%3 update); show details         epidemiological year:       07/2010-06/2011       date entere:       2012-02-15         disease:       invasive (unspecified/other)       date states:       2014-06-23         Sequence bin       275       N90:       6405         total length:       2195045 bp       N99:       3513         max length:       19869 bp       loci tagged:       1611         mean length:       782 top       detailed breakdown:       Display         N50:       34308        state states       state states         Schemes and loci       Image: plin browsing plin						(E-mail:			
epidemiological year: 072010-06/2011 date entered: 2012-02-15 disease: invasive (unspecified/other) date stamp: 2014-06-23 Sequence bin contigs: 275 N90: 6405 total length: 2195045 bp N95: 3513 max length: 109859 bp loci tagged: 1611 mean length: 7982 bp detailed breakdown: Display N50: 34308 Schemes and locl Schemes and locl Mali loci Gapsule Gapsule Mali select schemes within tree to display allele designations									
disease:       invasive (unspecified/other)       datestamp:       2014-06-23         Sequence bin       contigs:       275       N90:       6405         total length:       2195045 bp       N95:       3513         max length:       109859 bp       loci tagged:       1611         mean length:       7982 bp       detailed breakdown:       Display         N50:       34308       34308       Display         Schemes and loci       Navigate and select schemes within tree to display allele designations       Schemes information Processing         All loci       Genetic Information Processing       Navigate and select schemes within tree to display allele designations         Pilin       Pilin       Pilin       Pilin       Pilin         Pilin       Pilin       Typing       Pilin       Pilin	-								
Sequence bin contigs: 275 N90: 6405 total length: 2195045 bp N95: 3513 max length: 109859 bp loci tagged: 1611 mean length: 7982 bp detailed breakdown: Display N50: 34308 Schemes and loci All loci Capsule Capsule Capsule Capsule Pilin Pilin Pilin Pilin Typing									
contigs:       275       N90:       6405         total length:       2195045 bp       N95:       3513         max length:       109899 bp       loci tagged:       1611         mean length:       7982 bp       detailed breakdown:       Display         N50:       34308       Other controls       Schemes and loci         Schemes and loci         All loci       Capsule       Schemic Information Processing         Metabolism       Metabolism       Schemes within tree to display allele designations	disease:	invasive (unspecifi	ied/other)	uatestamp.	2014-00-23				
total length:     2195045 bp     N95:     3513       max length:     109859 bp     locitagged:     1611       mean length:     7982 bp     detailed breakdown:     Display       N50:     34308   Schemes and loci       All loci     Gapsule       Genetic Information Processing     Metabolism       Pilin     Pilin       Pilin     Typing	Sequence bin								
total length:     2195045 bp     N95:     3513       max length:     109859 bp     locitagged:     1611       mean length:     7982 bp     detailed breakdown:     Display       N50:     34308   Schemes and loci       All loci     Gapsule       Genetic Information Processing     Metabolism       Pilin     Pilin       Pilin     Typing	contine	275	N90-	6405					
mean length:     7982 bp     detailed breakdown:     Display       N50:     34308     34308   Schemes and locl       All loci     Image: Consule     Image: Consule									
N50: 34308 Schemes and locI  All loci Genetic Information Processing Height Genetic Information Processing H			loci tagged:	1611					
Schemes and loci			detailed breakdown:	Display					
All loci Capsule Genetic Information Processing Metabolism Pilin Typing	N50:	34308							
All loci Capsule Ca	Schemes and loci								
Capsule Capsul			Noviente enderste de la companya de						
Image: Construction Processing			Navigate and select schemes within tree	to display allele designations					
e- 💭 Pilin e- 🔊 Typing		mation Processing							
t- 🛄 Typing									
	1 1 3	ies							
Loci not in schemes									

Each record will have some or all of the following sections:

#### 10.1.1 Projects



This displays a list of projects that the isolate is a member of. Only projects that have a full description will be displayed.

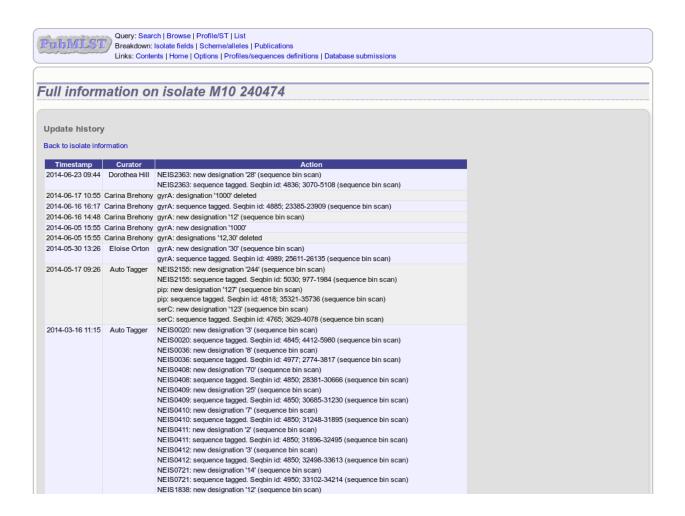
#### 10.1.2 Provenance metadata

Provenance/meta	data		
id:	18968	species:	Neisseria meningitidis
isolate:	M10 240474	serogroup:	В
strain designation:	B: P1.19-1,15-11: F3-9: ST-269 (cc269)	ENA accession:	ERR086224 → www.ebi.ac.uk
country:	UK	sender:	Dorothea Hill
continent:	Europe	curator:	Dorothea Hill, University of Oxford, UK (E-mail:
region:	South East		dorothea.hill@zoo.ox.ac.uk)
year:	2010	update history:	103 updates show details
epidemiological year:	07/2010-06/2011	date entered:	2012-02-15
disease:	invasive (unspecified/other)	datestamp:	2014-06-23

This section includes:

- provenance fields
- · housekeeping data
  - who sent the isolate
  - who last curated
  - record creation times
  - last update times
  - links to update history

The update link displays page with exact times of who and when updated the record.



#### **10.1.3 Publications**

Publications (8) show/hide

- Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007). Species status of Neisseria gonorrhoeae: evolutionary and epidemiological inferences from multilocus sequence typing. BMC Biol 5:35 [578 isolates]
- Didelot X, Urwin R, Maiden MC, Falush D (2009). Genealogical typing of Neisseria meningitidis. Microbiology 155:3176-86 93 isolates
- Jolley KA, Sun L, Moxon ER, Maiden MC (2004). Dam inactivation in Neisseria meningitidis: prevalence among diverse hyperinvasive lineages. BMC Microbiol 4:34 [24 isolates]
- Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005). The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in Neisseria meningitidis. Mol Biol Evol 22:562-9 [378 isolates]
- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Unvin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95:3140-5 107 isolates
- Thompson EA, Feavers IM, Maiden MC (2003). Antigenic diversity of meningococcal enterobactin receptor FetA, a vaccine component. Microbiology 149:1849-58
   Torisolates
- Unvin R, Russell JE, Thompson EA, Holmes EC, Feavers IM, Maiden MC (2004). Distribution of surface protein variants among hyperinvasive meningococci: implications for vaccine design. Infect Immun 72:5955-62 [78 isolates]

Wang JF, Caugant DA, Li X, Hu X, Poolman JT, Crowe BA, Achtman M (1992). Clonal and antigenic analysis of serogroup A Neisseria meningitidis with particular reference to epidemiological features of epidemic meningitis in the People's Republic of China. Infect Immun 60:5267-82 [47] isolates

This section includes full citation for papers linked to the isolate record. Each citation has a button that will return a dataset of all isolates linked to the paper.

If there are five or more references they will be hidden by default to avoid cluttering the page too much. Click the 'Show/hide' button to display them in this case.

#### 10.1.4 Sequence bin summary

Sequence bin			
contigs:	275	N90:	6405
total length:	2195045 bp	N95:	3513
max length:	109859 bp	loci tagged:	1611
mean length:	7982 bp	detailed breakdown:	Display
N50:	34308		

This section contains basic statistics describing the sequence bin. Clicking the 'Display' button navigates to the *sequence bin record*.

#### 10.1.5 Scheme and locus data

A hierarchical tree displays available schemes. Click within internal nodes to expand them.

Schemes and loci	
🛓 🕼 All loci	Navigate and select schemes within tree to display allele designations
- 🚛 Capsule	
- 💭 Genetic Information Processing	
🗁 🌆 Metabolism	
🗁 🌆 Pilin	
🖙 🚛 Typing	
- 💭 Other schemes	
🜆 Loci not in schemes	

Clicking any terminal node will display data available for a scheme or group of schemes.

Schemes and loci							
Genetic Information Processing			М	LST			
Pilin Typing MLST 	 adk aroE 0 S 15 S	fumC 9S	gdh 8 S	pdhC 11 S	pgm 9 S	<b>ST</b> 269	clonal complex ST-269 complex

Click an allele number within the scheme profile, will display the appropriate *allele definition record*. Clicking the green 'S' link will display the appropriate *sequence tag record*.

### **10.2 Allele definition records**

An allele definition record displays information about a defined allele in a sequence definition database.

-	
	ery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query wnload: Alleles   MLST profiles
	windad Andres (mice) profiles
Allele informa	ntion - abcZ: 2
Provenance/meta d	Jata
locus: a	abcZ i
allele: 2	2
	ITIGATACCS TIGCCGAAGG TITGGCGAAG ATTCGCGAIT TATTCGCCC TIACCACCGC GICGGCCAIG AGTIGGAAA CGGITCGGGT GAGGCTITGI TGAAGAACT CAACTGAATA CAACTGAAA TCGAACGGAA
	SACSECTEG AASCTEGATE CESCASTCAA SCAGACTIG GEGENATICE GETTECCEGA AAACGAAAAA ACGECAACC TITCCEGCEGE TAGAAAAAA CESTECECCI TESCELASEC ITEGETECLE AASCCEAACE TAITECTEGT EGACGAACCE ACCAACCAI TEGEATATCEA CECCEGAAA ATCECEGAAA ATCECTCAA ASCETITGAA GECAECTEG ITETEATATAC CACCACCEC CETTITITEG AAATATACE CACCEGEGAI
	STCGRACTCG ATC
length: 4	433
status: S	Sanger trace checked
date entered: 2	
datestamp: 2	2009-11-11
	Keith Jolley, University of Oxford, UK
curator: N	Man-Suen Chan, University of Oxford (E-mail: man-suen.chan@paediatrics.ox.ac.uk)
Profiles containing	this allele
MIST (	993 profiles
MESI.	aooyuuna
Isolate databases	
PubMI ST isolates: C	Contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles database there is at least one corresponding isolate
	deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample. 5041 isolates

If the allele is a member of a scheme profile, e.g. MLST, this will be listed. In this case, there will be a button to display all profiles of that scheme that contain the allele.

Similarly, if a *client database* has been setup for the database and the allele has been identified in an isolate, there will be a button to display all isolates that have that allele.

### 10.3 Sequence tag records

EIS0346 all	ele sequence:	d-18968									
ontig position											
sequence bin id:	4956	le	ngth:	744							
contig length:	22993	orient	ation:	forward							
	13864		plete:	•							
end:	14607	me	thod:	Illumina							
equence											
AIGTGGTIG AAGCGG GIGIIGGAA ACAACC AIGGIITAC GGGCAA GCAAAIGCG CIGGAA GCCCGGCIG IIGCGC	GGAG CCGTCGTCCT GCCGT GCCG TACATACGCG CAAACC CTGT TCCACGGAAA AGCGG CAAA CCGCCCGCGA ACTCG	GREG GAMAMACTAT TACCETACAS CC CGCA ACATTGCCEG AMAGCCCGAC AC CCEC CCTEGAAACCC AMAGCCCTT ET TETC CIGCIGGACA ATATACTCAM AG CGGC IGGCGGCAC CCGCCGTCCT ET CGCC TGGCGGACGC TCAMGCAGCT CA TCTC CCTC	AAATCC CCGGCG CCGTAG TCCTCG	GC CCGCAAAAGC TT TCAGACGGCA GG CTCGATGCCG GA CAGGCGTTTG	AAACCGTCCT TCGCCCCCGT CCTATGTCCA TCAAACCGGA	CAGCATTCCG TCCCGCCGCT CAAAACCTGT ACGGCAGGCG	CAGCGTCCGT TCGGGCATAA TGGGTGAAAA ATGATTGAAA	CCGAACAGCA CCAAACTTGC CCGCCGCCGT CTTTGTGCGC	TACCGGTCAG CGTTGTCAGC CGGCAACCCG CGGCCGCCCC	GCACGGCTCA CTGTGTCCGC ATGCCGTCTG TTCTTCATCA	AAACCATGAA CGACCGAGGA AAGCGGCAAT TCGACCATCC
ranslation											
T V L L Y 1 CACTGTAN 	R G A R L Q H ( A A H G F S I CGCGGCGCACGGCTICAGCAT: 	R A G G K T I T V C G Q A E K L L P Y A G R R K N Y Y R T GCGGCAGGCGGAAGAACTATTACCTATA 	SR AD AGCCGA -I EA KR KR	R N R R G K T E CCGGTAAAACCGAA : G S R R G S R R G A V V G E P S S	CRLN GATGCCGTCTTA I: PAVR LPSA CRP(	ATGGAGAAAA  : N I A G T L P I Q H C R	F1 E F2 F3				
K P I S P	T Q I R P Q K	: : : . N R P Q H S A A S Q T V L S I P Q R P K P S S A F R S V F	V R S E	T A Y R S Q H T G	GTAÇ QARL	N H E					

A sequence tag record displays information about the location within a contig of a region associated with a locus. The nucleotide sequence will be displayed along with upstream and downstream flanking sequence. The length of these flanking sequences can be modified within the *general options*.

If the tag is for a DNA locus and it is marked as a coding sequence, the three-frame translation will also be displayed.

### **10.4 Profile records**

(PubMLST) D	tuery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query ownload: Alleles   MLST profiles inks: Contents   Home   PorA   FetA   Options   Isolate Database
Profile inform	nation for ST-11 (MLST)
ST abcZ adk a	aros         fumC         gdh         pdhC         pgm         clonal complex           4         3         8         4         6         ST-11 complex/ET-37 complex
sender:	Paula Kriz, Paula Kriz and Keith Jolley
curator:	Keith Jolley, University of Oxford, UK (E-mail: keith.jolley@zoo.ox.ac.uk)
date entered:	2001-02-07
datestamp:	2009-11-11
Client database	
PubMLST isolates:	Contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample. 2899 isolates

A profile record displays information about a scheme, e.g. MLST, profile. Each allele number within the profile will be hyperlinked. Clicking these will take you to the appropriate *allele definition record*.

If a *client database* has been setup for the database and an isolate has the profile, there will be a button to display all isolates that have the profile.

## 10.5 Sequence bin records

-surrent mil	Links: Conte	Isolate fields   Scheme/ ents   Home   Options   P	Profiles/sequences de	innuons   Database sub	missions						
											Toggle
equen	ce bin for N	/10 240474									
ontig sum	nmary statistics		Contig size	distribution	Cumulative cont	tig lengt	h				
<ul> <li>Total le</li> <li>Minimu</li> <li>Maximu</li> <li>Mean l</li> <li>σ lengt</li> <li>N50: 3</li> <li>N90: 6</li> <li>N95: 3</li> </ul>	6405		Click to enlarg	je oad lengths							
	load sequences (FAS load sequences with	STA format) annotations (EMBL form		oudrongino							
• Downl	load sequences with	annotations (EMBL form	mat)	-	Locus	Start	End	Direction	EMBL format	Artemis 🖬	
• Downl		annotations (EMBL forn Original de	mat) esignation	Length Comments	Locus NEIS1151	Start 488	End		EMBL format		
• Downl	load sequences with Sequencing method	annotations (EMBL form	mat) esignation	Length Comments				Direction	EMBL format	Artemis i Artemis	
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151	488	1144	← ←			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150	488 1141	1144 2913	←			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149	488 1141 3155	1144 2913 3733	$\begin{array}{c} \leftarrow \\ \leftarrow \\ \rightarrow \end{array}$			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148	488 1141 3155 3896	1144 2913 3733 4513	$\begin{array}{c} \downarrow \\ \downarrow \\ \uparrow \\ \uparrow \end{array}$			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148 NEIS1147	488 1141 3155 3896 4521	1144 2913 3733 4513 5384	1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 +			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148 NEIS1147 NEIS1146	488 1141 3155 3896 4521 5397	1144 2913 3733 4513 5384 5831	1 + + + 1 1 + + + 1			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148 NEIS1147 NEIS1146 NEIS1145	488 1141 3155 3896 4521 5397 6141	1144 2913 3733 4513 5384 5831 6869	$\downarrow \downarrow \uparrow \uparrow \uparrow \downarrow \downarrow \uparrow$			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148 NEIS1147 NEIS1146 NEIS1145 NEIS1144 NEIS1143 NEIS1142	488 1141 3155 3896 4521 5397 6141 6892 8005 9288	1144 2913 3733 4513 5384 5831 6869 7923 8313 10541	1 + + + + + + + + + + + + + + + + + + +			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148 NEIS1147 NEIS1146 NEIS1145 NEIS1144 NEIS1143	488 1141 3155 3896 4521 5397 6141 6892 8005 9288	1144 2913 3733 4513 5384 5831 6869 7923 8313	11111111			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148 NEIS1147 NEIS1146 NEIS1145 NEIS1144 NEIS1143 NEIS1142	488 1141 3155 3896 4521 5397 6141 6892 8005 9288	1144 2913 3733 4513 5384 5831 6869 7923 8313 10541 12241	11111111			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148 NEIS1147 NEIS1146 NEIS1145 NEIS1145 NEIS1144 NEIS1142 NEIS1142 NEIS1140	488 1141 3155 3896 4521 5397 6141 6892 8005 9288 10613	1144 2913 3733 4513 5384 5831 6869 7923 8313 10541 12241 12744	111111111			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1148 NEIS1146 NEIS1146 NEIS1145 NEIS1144 NEIS1143 NEIS1142 NEIS1140 NEIS1139	488 1141 3155 3896 4521 5397 6141 6892 8005 9288 10613 12391 12824	1144 2913 3733 4513 5384 5831 6869 7923 8313 10541 12241 12744	↓ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↓ ↓ ↓ ↓ ↓			
• Downl	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1150 NEIS1149 NEIS1149 NEIS1146 NEIS1146 NEIS1145 NEIS1143 NEIS1143 NEIS1142 NEIS1140 NEIS1139 NEIS1138	488 1141 3155 3896 4521 5397 6141 6892 8005 9288 10613 12391 12824	1144 2913 3733 4513 5384 5831 6869 7923 8313 10541 12241 12241 122744 14362 15046	4 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			
Downl     equence	load sequences with Sequencing method	annotations (EMBL forn Original de	mat) esignation	Length Comments	NEIS1151 NEIS1149 NEIS1149 NEIS1148 NEIS1147 NEIS1145 NEIS1144 NEIS1144 NEIS1144 NEIS1142 NEIS1140 NEIS1137	488 1141 3155 3896 4521 5397 6141 6892 8005 9288 10613 12391 12824 14426 15102	1144 2913 3733 4513 5384 5831 6869 7923 8313 10541 12241 12241 122744 14362 15046	L L L T T T T L T L L L T T			

A sequence bin record contains information about that contigs associated with an isolate record. This includes:

- Number of contigs
- Total length
- Minimum length
- Maximum length
- N50, N90 and N95 values
- Size distribution charts

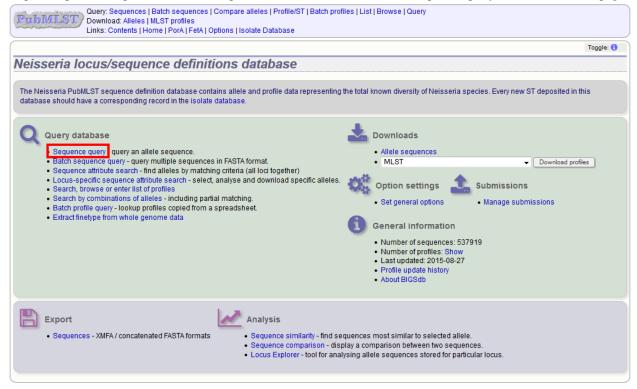
There are also links to download the contigs in FASTA or EMBL format.

Finally there is a table that shows the loci that are tagged on each contig. Individual contigs can also be downloaded in EMBL format.

### **Querying data**

# 11.1 Querying sequences to determine allele identity

Sequence queries are performed in the sequence definition database. Click 'Sequence query' from the contents page.



Paste your sequence in to the box - there is no need to trim. Normally, you can leave the locus setting on 'All loci' - the software should identify the correct locus based on your sequence. Sometimes, it may be quicker, however, to select the specific locus or scheme (e.g. MLST) that a locus belongs to.

**Note:** If the locus you are querying is a shorter version of another, e.g. an MLST fragment of a gene where the full length gene is also defined, you will need to select the specific locus or the scheme from the dropdown box. Leaving the selection on 'All loci' will return a match to the longer sequence in preference to the shorter one.

Click 'Submit'.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch p Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	vrofiles   List   Browse   Query
Sequence query - Neisseria locus/sequence definitio	ns
Please paste in your sequence to query against the database. Query sequences will be checked in partial matches will be identified if an exact match is notfound. You can query using either DNA or please select locus/schemeOrder results by	first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest peptide sequences. []  Alternatively upload FASTA file Action Select FASTA file: Browse_No file selected.

If an exact match is found, this will be indicated along with the start position of the locus within your sequence.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Brow Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	vse   Query
Sequence query - Neisseria locus/sequence definitions	
GACGCGGTGCGCGATGAAAAAGTCAAAGTCATCAAGTCATTGAAGCCGCTGACCGTCGAA	
1 exact match found. Translate query	
Allele         Length         Start position         End position         Flags         Comments           aroE: 8         490         136         625	

If only a partial match is found, the most similar allele is identified along with any nucleotide differences. The varying nucleotide positions are numbered both relative to the pasted in sequence and to the reference sequence. The start position of the locus within your sequence is also indicated.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Sequence query - Neisseria locus/sequence definitions	
Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - the partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences. []         — Please select locus/scheme       — Order results by –         MLST       Iocus         — Enter query sequence (single or multiple contigs up to whole genome in size)       — Alternatively upload FASTA file         GACCCCGTGCGCGTGAAAAAAGTCAAAGTCATCAAGTCATTGAAGCCGCTGACCGTCGAA       Select FASTA file:	ey do not need to be trimmed. The nearest Action Reset Submit
TCTSTCAATGAAAAATGTCGTGCCCGGGCGAATATATCCGCCGSCAAAGGCATGAACGGCTAT CTTGAAGAAATCAACTATCGGTTIGGCCAACGACATCACGCAGGTCAAAAGCATGACGACTAT TGAAGGACAACCGCCCTGCCGACGGCGGGCGGGGGGGGGG	
Translate query Closest match: aroE: 8 Show alignment	
Differences 2 differences found. []	
$^{259}\text{T}_{\rightarrow}^{393}\text{A}$ $^{468}\text{G}$ The locus start point is at position 136 of your query sequence. $\fbox$	

As an alternative to pasting a sequence in to the box, you can also choose to upload sequences in FASTA format by clicking the file browse button.

Query:         Sequences         Batch         sequences         Compare alleles         Profile/ST         Batch         profiles         L           Download:         Alleles         MLST         profiles         L         Links:         Contents         Home         PorA         FetA         Options         Isolate         Database	List   Browse   Query	
Sequence query - Neisseria locus/sequence definitions         Please paste in your sequence to query against the database. Query sequences will be checked first for ar partial matches will be identified if an exact match is not found. You can query using either DNA or peptide s         Please select locus/scheme       — Order results by –         All loci       Iocus         Enter query sequence (single or multiple contigs up to whole genome in size)		est

#### 11.1.1 Querying whole genome data

The sequence query is not limited to single genes. You can also paste or upload whole genomes - these can be in multiple contigs. If you select a specific scheme from the dropdown box, all loci belonging to that scheme will be checked (although only exact matches are reported for a locus if one of the other loci has an exact match). If all loci are matched, scheme fields will also be returned if these are defined. This, for example, enables you to identify the MLST sequence type of a genome in one step.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database								
Sequence query - Neisseria locus/sequence definitions								
Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The neare partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences.	st							
Please select locus/scheme								
MLST v locus v								
- Enter query sequence (single or multiple contigs up to whole genome in size) - Alternatively upload FASTA file - Action - Actio								
>4758   NODE_192_length_1326_cov_47.828808 Select FASTA file: Reset Submit								
TAGAACAACAGCAATATTCAAAGATTATCTGAAAGTCCGAGATTCTAGATTCCCGCTTTC GCGGGAATGACGAAAAGCAAGCCGTAGGTCGGATACTTGTATCCGACAAAAGCCTGCCAT								
CICAAAIAGCCGICGGAIICCGACAGAAICCGACCIGCCAAACCGGGCGGG								
SECAETIASTACSCAARICSAACACACACACACAAAAASCCCCSATICSSATITCCAAT CSGSCTTTTTTCSCCCCGTTTTSTCCCCGTTAAAAAATACCCCATACAAAAATAASTG -								
AATTAACAAAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAGATAGTACGGTAAGG								
7 exact matches found.								
Allele Length Start position End position Flags Comments								
adk: 10 465 938327 938791								
aroE: 15 490 1775325 1775814 a								
fumC:9 465 1134240 1134704								
gdh: 8 501 961439 961939								
pdhC:11 480 1341678 1342157								
pgm: 9 450 1416246 1416695								
MLST								
ST 269								
clonal complex ST-269 complex								

### 11.2 Searching for specific allele definitions

There are two query pages available that allow searching for specific allele definitions. The first allows querying of all loci together by criteria that are common to all. The second is a locus-specific attribute query that can search on any extended attributes that may be defined for a locus. This locus- specific query also allows you to paste in lists of alleles for download or analysis.

#### 11.2.1 General (all loci) sequence attribute search

To retrieve specific allele designations, click 'Sequence attribute search' on a sequence definition database contents page.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	iles   List   Browse   Query
	Toggle: 🚯
Neisseria locus/sequence definitions database	
The Neisseria PubMLST sequence definition database contains allele and profile data representing database should have a corresponding record in the isolate database.	the total known diversity of Neisseria species. Every new ST deposited in this
Query database           • Sequence query - query an allele sequence.           • Batch sequence query - query multiple sequences in FASTA format.           • Sequence attribute search find alleles by matching criteria (all loci together)           • Locus-specific sequence attribute search - select, analyse and download specific alleles.           • Search, browse or enter list of profiles           • Search by combinations of alleles - including partial matching.           • Batch profile query - lookup profiles copied from a spreadsheet.           • Extract finetype from whole genome data	<ul> <li>Downloads         <ul> <li>Allele sequences</li> <li>MLST</li> <li>Download profiles</li> </ul> </li> <li>Option settings         <ul> <li>Submissions</li> <li>Set general options</li> <li>Set general options</li> <li>Manage submissions</li> </ul> </li> <li>General information         <ul> <li>Number of sequences: 537919</li> <li>Number of profiles: Show</li> <li>Last updated: 2015-08-27</li> <li>Profile update history</li> <li>About BIGSdb</li> </ul> </li> </ul>
Sequence comparison - disp	quences most similar to selected allele. Ilay a comparison between two sequences. ysing allele sequences stored for particular locus.

Enter your query using the dropdown search box - additional terms can be added by clicking the '+' button.

Designations can be queried using standard operators.

Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database
Toggle: i
Query sequences for Neisseria locus/sequence definitions database
Some loci have additional fields which are not searchable from this general page. Search for these at the locus-specific query page. Use this page also for access to the sequence analysis or export plugins.
Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a locus that uses integer allele ids using the drop-down list.
Please enter your search criteria below (or leave blank and submit to return all records). Matching sequences will be returned and you will then be able to update their display and query settings.
Search criteria Display
locus v = v abcZ + 1 Order by: locus v ascending v
Display: 25 V records per page []
> Filter query by
Reset Submit

Click submit.

Cuery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database
Toggle: [i]
uery sequences for Neisseria locus/sequence definitions database
Some loci have additional fields which are not searchable from this general page. Search for these at the <u>locus-specific query</u> page. Use this page also for access to the sequence analysis or export source that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a bocus that uses integer allele ids using the drop-down list.
Please enter your search criteria below (or leave blank and submit to return all records). Matching sequences will be returned and you will then be able to update their display and query settings.
Combine searches with: AND - Order by: locus - ascending -
locus v = v abcZ + i Display: 25 v records per page i allele id v = v 5
Filter query by Action Reset Submit
record returned. Click the hyperlink for detailed information.
locus allele id sequence sequence length comments flags abc2 5 TITGATACCETTECC TOETCEAACTOGATC 433

#### Click the hyperlinked results to display allele records.

PubMLST Dowr	ry: Sequences   Batch sequences   Compar nload: Alleles   MLST profiles s: Contents   Home   PorA   FetA   Options   I		rofiles   List   Browse   Query	
				Toggle: i
Query sequend	ces for Neisseria locus/	sequence defin	itions database	
plugins. Also note that some loci ir	-		·	this page also for access to the sequence analysis or export ally rather than numerically unless you filter your search to a
	criteria below (or leave blank and submit to		sequences will be returned and you wi	II then be able to update their display and query settings.
		— Display		
Combine searches with		Order by:		
	→ = → abcZ	+ i Display:	25 👻 records per page 👔	
allele id 🗸 🗸	- = • 5			
Filter query by	Action Reset Submit			
1 record returned. Click the	e hyperlink for detailed information.			
locus allele id abcZ 5 TTTGATA		nce length comments flags 433		

	ownload: Alleles   MLST profiles inks: Contents   Home   PorA   FetA   Options   Isolate Database
llele inform	ation - abcZ: 5
Provenance/meta	data
locus:	abcZ i
allele:	5
sequences:	TITGATACCS TIGCCGAAGS TITGGGCGAA ATICGCGATI TAITGCGCCS ITAICATCAI GICAGCCAIG AGTIGGAAAA TGGTICGAGI GAGGCITIGI TGAAAGAGCI TAACGAATIG CAACITGAAA TCGAAGCGAA GGACGGCIGG AAGCCIGGAIG CGGCGAGICAA GCAGACITIG GGICGAACIIG GITTGCCAGA AAACGAAAAA AICGGCAACC CGIGIGIGCC TAGCGCAGGC TIGGGIGCGG AAGCCIGAIG TAITGGIGCGGAACGG GACGGAACCG ACCAACCAIT IGGACATIGA CGCGAITAIT IGGCIGGAAA AICIGCITAA AGCGITIGAA GGCAGCCIGG IIGGAITAC CCACGACCGC CGITITIIGG ACAATAICGC CACCGCAIC GICGAACCG AIC
length:	433
status:	Sanger trace checked
date entered:	2001-02-07
datestamp:	2009-11-11
	Keith Jolley, University of Oxford, UK
curator:	Man-Suen Chan, University of Oxford (E-mail: man-suen.chan@paediatrics.ox.ac.uk)
Profiles containin	g this allele
MLST:	183 profiles
Isolate databases	
PubMLST isolates:	Contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Various search criteria can also be selected by combining with filters. Click the filter heading to display these.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles   List   Browse   Query Download: Alleles   MLST profiles   List   Contents   Home   PorA   FetA   Options   Isolate Database
Toggle: [
Query sequences for Neisseria locus/sequence definitions database
Some loci have additional fields which are not searchable from this general page. Search for these at the locus-specific guery page. Use this page also for access to the sequence analysis or export
plugins.
Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a locus that uses integer allele ids using the drop-down list.
Please enter your search criteria below (or leave blank and submit to return all records). Matching sequences will be returned and you will then be able to update their display and query settings.
Search criteria
allele id • < • 10 + ; Order by: locus • ascending •
Display: 25 → records per page []
locus: abcZ T i Reset Submit
status:
sender:
curator:
allele flag:
9 records returned. Click the hyperlinks for detailed information.
locus allele id sequence sequence length comments flags
abcz 1 TITGATACTGTIGCC TIGICGAACICGATC 433
abcz 2 TITGATACCGITGCC TIGICGAACTCGAIC 433
abcz 3 TITGATACCGTIGCC TIGTIGAACTIGACC 433
abcz 4 TITGATACCGTIGCC TIGTCGAACICGAIC 433
abcz 5 ITTGATACCGTTGCC TCGTCGAACTCGATC 433
abcz 6 ITTGATACCGTTGCC TTGTCGAACCCGATC 433
abcz 7 IIIIGAIACIGIIGCC IIGICGAACICGAIC 433
abcz 8 TITGATACCGTIGCC TIGTCGAACTIGACC 433
abcZ 9 TITGATACCGTIGCC TIGICGAACTCGATC 433

#### 11.2.2 Locus-specific sequence attribute search

Some loci have *extended attribute fields*. To query these, click 'Locus-specific sequence attribute search' on a sequence definition database contents page.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	iles   List   Browse   Query
	Toggle: 🕄
Neisseria locus/sequence definitions database	the total known diversity of Neisseria species. Every new ST deposited in this
Query database         • Sequence query - query an allele sequence.         • Batch sequence query - query multiple sequences in FASTA format.         • Sequence attribute search - find alleles by matching criteria (all loci together)         • Locus-specific sequence attribute search         • Search of combinations of alleles - including partial matching.         • Batch profile query - lookup profiles copied from a spreadsheet.         • Extract finetype from whole genome data	<ul> <li>Downloads         <ul> <li>Allele sequences</li> <li>MLST</li> <li>Download profiles</li> </ul> </li> <li>Option settings         <ul> <li>Set general options</li> <li>Set general options</li> <li>Set general options</li> <li>Manage submissions</li> </ul> </li> <li>General information         <ul> <li>Number of sequences: 537919</li> <li>Number of profiles: Show</li> <li>Last updated: 2015-08-27</li> <li>Profile update history</li> <li>About BIGSdb</li> </ul> </li> </ul>
Sequence comparison - disp	quences most similar to selected allele. Jay a comparison between two sequences. ysing allele sequences stored for particular locus.

Pick the required locus from the dropdown box.

Cuery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Help 🗹	Toggle: 🚺
Query PorA VR2 sequences - Neisseria locus/sequence definitions database	
Locus PorA_VR2 Page will reload when changed  • Further information is available for this locus. Please enter your search criteria below (or leave blank and submit to return all records).	Modify form options
Allele fields Display Action	

The fields specific for that locus will be added to the dropdown query boxes.

	LST/ c	Query: Sequences   Batch s Download: Alleles   MLST pr Links: Contents   Home   Po	ofiles	alleles   Profile/ST   solate Database	Batch pro	ofiles   I	List   Browse	Query					
											He	elp 🗗	Toggle: 🚯
Query H	PorA	VR2 sequence	s - Neisseri	a locus/seq	quen	ce d	efinitio	ons data	base				
Locus: Por	A \/D2		<ul> <li>Page will reload</li> </ul>										
	-			when changed									Modif
<ul> <li>Furth</li> </ul>	ner inform	ation is available for this loo	cus.										form option
Please ente	r your sea	arch criteria below (or leave	blank and submit to i	return all records).									
Allele fie	alde			Disp	Jov					n			
	eius	- 2				-llele i	1	a second in a		-	_		
family		• = • 2			rder by:			ascending	<ul> <li>Rese</li> </ul>	t Sub	mit		
					)isplay:	25 👻	records pe	r page 🕦					
_		- 25 displayed). Click the h	yperlinks for detailed	information.									
Page: 1	2 3	4 > Last											_
age: 1	2 3 allele id	4 > Last sequence	sequence length	information. comments	_	varian	t  old name		mAb1 reactivit		nAb2 reac	tivity flag	s
age: 1 locus a PorA VR2	2 3 allele id 2	4 > Last Sequence HFVQQTPKSQPTLVP	sequence length		2			MN16C13F4	+	AF202	-	tivity flag	IS
age: 1 <b>locus</b> a PorA VR2 PorA VR2	2 3 allele id 2 2-1	4 > Last Sequence HFVQQTPKSQPTLVP HFVQQPPKSQPTLVP	sequence length 15 15		2	1	t <mark>old name</mark> 2b				nAb2 reac - +	tivity flag	IS
age: 1 locus a PorA VR2 PorA VR2 PorA VR2	2 3 allele id 2 2-1 2-10	4 > Last Sequence HFVQQTPKSQPTLVP HFVQQPPKSQPTLVP HFVQQAPQSQSTLVP	sequence length 15 15 15 15		2 2 2	1 10		MN16C13F4	+	AF202	-	tivity flag	ß
locus PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2	2 3 allele id 2 2-1 2-10 2-11	4 > Last Sequence HFVQQTPKSQPTLVP HFVQQPPKSQPTLVP HFVQQAPQSQSTLVP HFVLQTPQSQPTLVP	sequence length 15 15 15 15 15		2 2 2 2	1 10 11		MN16C13F4	+	AF202	-	tivity flag	B
age: 1 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2	2 3 allele id 2 2-1 2-10 2-11 2-12	4 > Last Sequence HFVQTPKSQPTLVP HFVQQPPKSQPTLVP HFVQQAPQSQSTLVP HFVQQIPKSQPTLVP HFVQQIPKSQPTLVP	sequence length 15 15 15 15 15 15 15		2 2 2 2 2 2	1 10 11 12		MN16C13F4 MN16C13F4	+	AF202	-	tivity flag	B
age: 1 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2	2 3 allele id 2 2-1 2-10 2-11 2-12 2-13	4 > Last Sequence HFVQQTPKSQPTLVP HFVQQPKSQPTLVP HFVQQAPQSQPTLVP HFVQQTPKSQPTLVP YFVQQTPQSQPTLVP	sequence length 15 15 15 15 15 15		2 2 2 2 2 2 2 2	1 10 11 12 13	2b	MN16C13F4	+	AF202	-	tivity   flag	2
age: 1 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2	2 3 allele id 2 2-1 2-10 2-11 2-12 2-13 2-14	4 > Last Sequence HFVQQTPKSQPTLVP HFVQQPKSQPTLVP HFVQQPKSQPTLVP HFVQQTPQSQPTLVP HFVQQTPQSQPTLVP HFVQQTPQSQPTLVP HFVQQKLASKPTLVP	sequence length 15 15 15 15 15 15 15 15		2 2 2 2 2 2 2 2 2 2 2	1 10 11 12 13 14	2b 33	MN16C13F4 MN16C13F4	+	AF202	-	tivity   flag	8
age: 1 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2 PorA VR2	2 3 allele id 2 2-1 2-10 2-11 2-12 2-13 2-14 2-15	4 > Last Sequence HFVQQTPKSQPTLVP HFVQQPPKSQPTLVP HFVQQPQSQSTLVP HFVQQTPQSQPTLVP HFVQQTPQSQPTLVP HFVQQKSISKPTLVP HFVQQKSISKPTLVP	sequence length 15 15 15 15 15 15 15 15 15		2 2 2 2 2 2 2 2 2 2 2 2 2	1 10 11 12 13 14 15	2b 33 33a (33-1)	MN16C13F4 MN16C13F4	+	AF202	-	tivity   flag	5
Page: 1	2 3 allele id 2 2-1 2-10 2-11 2-12 2-13 2-14 2-15 2-16	4 > Last Sequence HFVQQTPKSQPTLVP HFVQQPKSQPTLVP HFVQQPKSQPTLVP HFVQQTPQSQPTLVP HFVQQTPQSQPTLVP HFVQQTPQSQPTLVP HFVQQKLASKPTLVP	sequence length 15 15 15 15 15 15 15		2 2 2 2 2 2 2 2 2 2 2	1 10 11 12 13 14	2b 33	MN16C13F4 MN16C13F4	+	AF202	-	tivity   flag	IS IS IS IS IS IS IS IS IS IS IS IS IS I

#### The query form can be modified by clicking the 'Modify form options' tab:

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Help 🗹	Toggle: 🜖
Query abcZ sequences - Neisseria locus/sequence definitions database	
Locus: abcZ	Modify form options
Please enter your search criteria below (or leave blank and submit to return all records).	
Allele fields Display Action allele id • = • • • • • • • • • Order by: allele id • ascending • Reset Submit Display: 25 • records per page •	

A list box can be added by clicking the 'Show' button for 'Allele id list box'.

PubMLST	Download: Alleles Links: Contents	s   MLST profiles Home   PorA   FetA	es   Compare alleles   Profile/   Options   Isolate Database eria profile/sequ			s database	Help 🕫	Toggle: ()
	mation is available earch criteria belov • =	for this locus.	age will reload when changed nd submit to return all records	Display Order by:	allele id 25 👻 records	ascending s per page	X Modify form parameters Click to add or remove additional que • • • Allele fields • • • Allele id list box • • • Filters	Modify form option: ary terms:

Close the form modification tab and you can now enter a list of allele ids for retrieval.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Help 🖸	Toggle: 🜖
Query abcZ sequences - Neisseria locus/sequence definitions database	
Locus: abcZ	Modify form options
Please enter your search criteria below (or leave blank and submit to return all records).  Allele fields Allele id ist Display Order by: allele id  ascending  Display: 25  records per page  Action Reset Submit	

Various analysis and export options will be available for use on the retrieved sequences. These include FASTA output and *Locus Explorer* analysis.

Further information is available for this locus. Please enter your search criteria below (or leave blank and sui	hmit to return all rec	ords)		form options
Allele fields	+ 0	Allele id list	Display       Order by:     allele id     ✓ ascending       Display:     25     ✓ records per page	
5 records returned. Click the hyperlinks for detailed information.		5	Reset Submit	
abcZ       1       TITIGATACTGITGCC        TITGCGAACTCGATC         abcZ       2       TITGATACCGTTGCC        TITGCGAACTCGATC         abcZ       3       TITGATACCGTTGCC        TIGCGAACTCGATC         abcZ       4       TITGATACCGTTGCC        TIGCGAACTCGATC         abcZ       5       TITGATACCGTTGCC        TIGCGAACTCGATC         AbcZ       5       TITGATACCGTTGCC        TCGTCGAACTCGATC	sequence length of 433 433 433 433 433 433 433	comments flags		
Export: FASTA Table Analysis: Locus Explorer				

### 11.3 Browsing scheme profile definitions

If a sequence definition database has schemes defined that include a primary key field, i.e. collections of loci that together create profiles, e.g. for MLST, these can be browsed by clicking the link to 'Search, browse or enter list of profiles'.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profile Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	Toggle: ()							
Neisseria locus/sequence definitions database								
<ul> <li>have a corresponding record in the isolate database.</li> <li>Query database</li> <li>Sequence query - query an allele sequence.</li> <li>Batch sequence query - query multiple sequences in FASTA format.</li> <li>Sequence attribute search - find alleles by matching criteria (all loci together)</li> <li>Locus spacific sequence attribute search - select, analyse and download specific alleles.</li> <li>Search by combinations of alleles - including partial matching.</li> <li>Batch profile query - locy profiles composed from a spreadsheet.</li> <li>Extract finetype from whole genome data</li> </ul>	Downloads     Allele sequences     MLST     Option settings     Submissions     Set general options     Set general information     Number of sequences: 537919     Number of sequences: 537919     Number of profiles: Show     Last updated: 2015-08-27     Profile update history     About BIGSdb							
Sequence comparison - disp	quences most similar to selected allele. play a comparison between two sequences. lysing allele sequences stored for particular locus.							

Leave query form fields blank (the display of these may vary depending on modification options set by the user). Choose the field to order the results by, the number of results per page to display, and click 'Submit'.

Pul	Cuery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database								
	Toggle: 🚯								
Sea	Search or browse profiles - Neisseria locus/sequence definitions								
- cu									
Enter	Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.								
					0 010			an records, modily form param	Modif
		s/sch	eme fi	elds –					Display/sort options Action form option
S	Т			-	=		•		+ () Order by: ST - ascending - Reset Submit
									Display: 25 👻 records per page 🕕
Brows	ang a	anrec	oras.						
10056	reco	ords r	returne	ed (1 -	25 di	splaye	d). Cl	ick the hyperlinks for detailed in	nformation.
_		_	-						
Page:		2	3	4	5	6 7		9 > Last	
									_
	_	abcZ	aroE	fumC	gdh	pdhC			
1	3	1	1	1	1	1	3	ST-1 complex/subgroup I/II	
2	3	1	4	7	1	1	3	ST-1 complex/subgroup I/II	
3	3	1	1	1	1	23	13	ST-1 complex/subgroup I/II	
4	3	1	3	1	4	2	3	ST-4 complex/subgroup IV	
5	1	1	2	1	3	2	3	ST-5 complex/subgroup III	
6	1	1	2	1	3	2	11	ST-5 complex/subgroup III	
/	1	1	2	1	3	2	19	ST-5 complex/subgroup III	
8	3	2	7	2	8	5	2	ST-8 complex/Cluster A4	
-	3	2	8	10 2	8	5 15	2	ST-8 complex/Cluster A4	
	3	2		2	8	4	2	ST-8 complex/Cluster A4 ST-11 complex/ET-37 complex	M .
	3	2	4	3 16	8	4	20	SI-TI complexEI-37 complex	λ
12		4	15	7	0	11	20	ST 260 complex	

Clicking the hyperlink for any profile will display full information about the profile.

	Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Profile/ST / Download: Alleles   MLST profiles										
Lin	Links: Contents   Home   PorA   FetA   Options   Isolate Database										
	Help C*										
<b>Profile inform</b>	Profile information for ST-11 (MLST)										
ST adk abcZ a	roE         fumC         gdh         pdm         clonal complex           4         3         8         4         6         ST-11 complex/ET-37 complex										
	Paula Kriz, Paula Kriz and Keith Jolley Keith Jolley, University of Oxford, UK (E-mail: keith.jolley@zoo.ox.ac.uk)										
	2 updates show details										
date entered:											
datestamp:	2013-04-27										
Client database											
I	Contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample. 1133 isolates										

## **11.4 Querying scheme profile definitions**

Click the link to 'Search, browse of enter list of profiles'.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profile Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	îles   List   Browse   Query
	Toggle: 🜖
Neisseria locus/sequence definitions database	
The Neisseria PubMLST sequence definition database contains allele and profile data representing have a corresponding record in the isolate database.	the total known diversity of Neisseria species. Every new ST deposited in this database should
Query database           • Sequence query - query an allele sequence.           • Batch sequence query - query multiple sequences in FASTA format.           • Sequence attribute search - find alleles by matching criteria (all loci together)           • Locus spacific sequence attribute search - select, analyse and download specific alleles.           • Search by combinations of alleles - including partial matching.           • Batch profile query - lookup profiles copied from a spreadsheet.           • Extract finetype from whole genome data	<ul> <li>Downloads         <ul> <li>Allele sequences</li> <li>MLST</li> <li>Download profiles</li> </ul> </li> <li>Option settings         <ul> <li>Set general options</li> <li>Set general options</li> <li>Manage submissions</li> </ul> </li> <li>General information         <ul> <li>Number of sequences: 537919</li> <li>Number of profiles: Show</li> <li>Ast update(2015-08-27)</li> <li>Profile update history</li> <li>About BICSdb</li> </ul> </li> </ul>
Sequence comparison - dis	quences most similar to selected allele. play a comparison between two sequences. ysing allele sequences stored for particular locus.

Enter the search criteria you wish to search on. You can add search criteria by clicking the '+' button in the 'Locus/scheme fields' section. These can be combined using 'AND' or 'OR'.

Pub	Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database									
									Toggle: 🤅	
Sear	ch	or	bro	WS	e p	rofi	les	- Neisseria locu	s/sequence definitions	
Lo	cus/	schem	ne fielo	ds —			wsea	all records. Modify form paramet	ters to filter or enter a list of values.  Display/sort options  Action  opti	
dat	Combine searches with: AND date entered sender (surname) = Jolley									
5051 re Page:	_	s retur 2					Click	the hyperlinks for detailed inform	nation.	
ST 5001		abcZ 12	aroE 12	fumC 352	gdh 9	pdhC 18	pgm 9	cional complex		
5002	5	2	9	9	9	6	8	ST-41/44 complex/Lineage 3		
5003	5	9	6	143	5	119	18			
5004	8	7	185	26	10	1	16	ST-18 complex		
5005	8	7	10	19	10	351	16	ST-18 complex		
5006										
5007										
5008	7	12	381	91	5	21	16			
5009	8	187	10	116	10	15	20			
5010	8	7	10	17	10	1	9	ST-18 complex		
5011	3	7	72	26	10	1	16			

Each field can be queried using standard operators.

Clicking the hyperlink for any profile will display full information about the profile.

Publ	Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database													
	Toggle											Toggle: 🚯		
Sear	ch	or	bro	ws	e n	rofi	iles	- Neisseria Io	cus/s	equence	definitio	ns		
	Search or browse profiles - Neisseria locus/sequence definitions													
Enter se	arc	h crite	eria or	leave	blank	to bro	wse a	all records. Modify form par	rameters to	filter or enter a	listofvalues			
					Jun			in records, modify form par	annotoro to					Modify
Loo	us/s	schen	ne fielo	ds —						— Display/sor	t options		Action	form options
Comb	Combine searches with: AND - Order by: ST - ascending - Reset Submit								opuons					
date	date entered v > v 2013-02-01 + 0 Display: 25 v records per page 0													
sen	der (	(surna	ime)	•	-		-	Jolley						
5051 rec Page:							Click 1	he hyperlinks for detailed i	information	L.				
_					_									
ST	adik	ahc7	aroE	fumC	adh	pdhC	nam	clonal complex						
5001	6	12	12	352	9	18	9	cional complex	_					
5002	5	2	9	9	9	6	8	ST-41/44 complex/Lineag	e 3					
5003	5	9	6	143	5	119	18							
5004	8	7	185	26	10	1	16	ST-18 complex						
5005	8	7	10	19	10	351	16	ST-18 complex						
5006	8	13	10	19	10	1	9	ST-18 complex						
5007	8	7	10	3	9	15	20							

Other query options are available by clicking the 'Modify form options' tab.

Query:         Sequences         Batch sequences         Compare alleles         Profile/ST         Batch profiles         List         Browse         Query           Download:         Alleles         MLST profiles         Links:         Contents         Home         PortA         Options         Isolate         Database								
	Toggle: 🚺							
Search or browse profiles - Neisseria locus/sequence definitions								
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.								
Locus/scheme fields Display/sort options Action Action FT ascending v Reset	Submit Modify							
Display: 25  → records per page 0								

For example, you can enter a list of attributes to query on by clicking the 'Show' button next to 'Attribute values list'.

Pi	PTDMLST Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database										
Log	Logged in: Keith Jolley (keith). 🔂 Log out   Change password Toggle: 🚯										
Se	Search or browse profiles - Neisseria profile/sequence definitions										
En	iter search cri	iteria or leave	blank to bro	wse all rec	ords. Modify for	m parameters to fi	Iter or enter a	ist of values.	(	×	Modify
	-Locus/sche	eme fields —					-Display/sort	options ——			form
	ST	-	=	-		+ 0	Order by:	ST	<ul> <li>ascen</li> </ul>	Modify form parameters	options
							Display:	25 🔻 reco	rds per page	Click to add or remove additional que	ery terms:
										• 😑 Locus/scheme field v	alues
										Attribute values list	
										Filters	
									(		

A list box will appear within the page. Hide the form modification tab by clicking the 'X' in the corner or the purple tab again. Now you can choose the attribute to search on along with a list of values.

िया	Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database												
Τορ									Toggle: 🚯				
Search or browse profiles - Neisseria locus/sequence definitions													
	•												
Enter	sea	arch c	riteria	or lea	ve bla	ank to b	rows	e all records. Modify form paramet	ers t	o filter or enter a list of value	s.		Modify
—L	.ocu	s/sch	eme f	ields -						— Attribute values list —		Display/sort options	form
S	т				=			+	0	Field: ST 👻		Order by: ST - ascending	options
										1		Display: 25 v records per page 1	
										2			
										3			
										•		Reset Submit	
4 reco	rds	return	ed. C	lick th	e hype	erlinks	for de	etailed information.					
OT		7		5	N and D	- db C							
	аак 3	abcz 1	aroe 1	1	gan 1	pdhC 1	pgm 3	cional complex ST-1 complex/subgroup I/II					
2	3	1	4	7	1	1	3	ST-1 complex/subgroup I/II					
3	3	1	1	1	1	23		ST-1 complex/subgroup I/II					
4	3	1	3	1	4	2	3	ST-4 complex/subgroup IV					
Anal	vei	e tor	de:										
Ana	Analysis tools:												
Analy	sis:	BU	RST										
Exp	port	Se	quen	es									

List values will be combined with any other attributes entered in the query form allowing complex queries can be constructed.

You can also add filters to the form by again clicking the 'Modify form options' tab and selecting 'Filters'.

Query: Sequences   Batch sequences   Compare a Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Iso			Help IZ Toggle: 8						
Search or browse profiles - Neisseria p	profile/sequence definitions								
Enter search criteria or leave blank to browse all records. Modify form particular search criteria or leave blank to browse all records. Modify form particular search criteria search criteri	Arameters to filter or enter a list of values. Attribute values list Field: ST 1 2 3 4 	• 🕒 Locus/s	Additional query terms: cheme field values values list						

Available filters will vary depending on the database. These will be combined with other query criteria or lists of attributes.

Query: Sequences   Batch sequences   Compare a Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isol		Browse   Query
		Toggle: 🚺
Search or browse profiles - Neisseria lo	ocus/sequence defin	tions
Enter search criteria or leave blank to browse all records. Modify form pa	rameters to filter or enter a list of value	s. Modify
Locus/scheme fields	Attribute values list	Filters form
ST 🗸 = 🗸	+ 🚯 Field: ST 🗸	Publication: Options
	1 2 3 4	clonal complex: ST-4 complex/subgroup IV ▼ Display/sort options Action Order by: ST ← ascending ← Reset Submit Display: 25 ← records per page 
1 record returned. Click the hyperlink for detailed information.		
ST     adk     abcZ     aroc     fumC     gdh     pdhC     pgm     clonal complex       4     3     1     3     1     4     2     3     ST-4 complex/subgroup IV       Analysis tools:     Export     Sequences		

### 11.5 Investigating allele differences

#### 11.5.1 Sequence similarity

To find sequences most similar to a selected allele within a sequence definition database, click 'Sequence similarity' on the contents page.

Profit         Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database							
Neisseria locus/sequence definitions database							
The Neisseria PubMLST sequence definition database contains allele and profile data representing the this database should have a corresponding record in the isolate database.	e total known diversity of Neisseria species. Every new ST deposited in						
Query database           • Sequence query - query an allele sequence.           • Batch sequence query - query multiple sequences in FASTA format.           • Sequence attribute search - find alleles by matching criteria (all loci together)           • Locus-specific sequence attribute search - select, analyse and download specific alleles.           • Search, browse or enter list of MLST profiles           • Search by combinations of MLST alleles - including partial matching.           • Batch profile query - lookup MLST profiles copied from a spreadsheet.           • Extract finetype from whole genome data	<ul> <li>Downloads         <ul> <li>Allele sequences</li> <li>MLST profiles</li> </ul> <ul> <li>Set general options</li> <li>Set general options</li> <li>Manage submissions</li> </ul> </li> <li>General information         <ul> <li>Number of sequences: 124770</li> <li>Number of profiles (MLST): 10058</li> <li>Last updated: 2015-08-19</li> <li>Profile update history</li> <li>About BIGSdb</li> </ul> </li> </ul>						
Sequence comparison - displa	ences most similar to selected allele. y a comparison between two sequences. ing allele sequences stored for particular locus.						

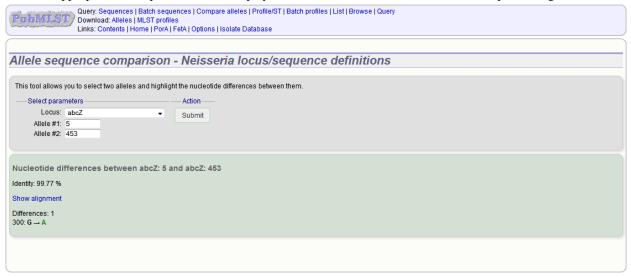
Enter the locus and allele identifier of the sequence to investigate and the number of nearest matches you'd like to see, then press submit.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database							
ind most similar alleles - Neisseria locus/sequence definitions							
This page allows you to find the most similar sequences to a selected allele using BLAST.							
Select parameters     Action       Locus:     abcZ        Allele:     5       Number of results:     10 -							

A list of nearest alleles will be displayed, along with the percentage identity and number of gaps between the sequences.

	PTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT									
ind m	oct ci	milor a			seria locus/sequenc	dofinitions				
mu m	031 311	iiiiai a	mere	:3 - Ne	seria locus/sequence	: uemnuons				
This page	allows vou t	o find the m	nost sim	nilar sequen	s to a selected allele using BLAST.					
	parameters				Action					
- Oelect		s: abcZ								
	Allel				Reset Submit					
Num	ber of result									
bo7 F										
abcZ-5										
Allele	% Identity				Compare					
Allele abcZ: 657	99.77	Mismatche: 1	s Gaps 0	433/433	Compare Compare abcZ: 657					
Allele										
Allele abcZ: 657	99.77	1	0	433/433	Compare abcZ: 657					
Allele abcZ: 657 abcZ: 453	99.77 99.77	1 1	0	433/433 433/433	Compare abcZ: 657 Compare abcZ: 453					
Allele abcZ: 657 abcZ: 453 abcZ: 405	99.77 99.77 99.77	1 1 1	0 0 0	433/433 433/433 433/433	Compare abcZ: 657 Compare abcZ: 453 Compare abcZ: 405					
Allele abcZ: 657 abcZ: 453 abcZ: 405 abcZ: 404	99.77 99.77 99.77 99.77 99.77	1 1 1 1	0 0 0 0 0	433/433 433/433 433/433 433/433	Compare abc2: 657 Compare abc2: 453 Compare abc2: 405 Compare abc2: 404 Compare abc2: 213					
Allele abcZ: 657 abcZ: 453 abcZ: 405 abcZ: 404 abcZ: 213	99.77 99.77 99.77 99.77 99.77 99.77	1 1 1 1 1	0 0 0 0 0	433/433 433/433 433/433 433/433 433/433	Compare abc2: 657 Compare abc2: 453 Compare abc2: 405 Compare abc2: 404 Compare abc2: 213 Compare abc2: 166					
Allele abcZ: 657 abcZ: 453 abcZ: 405 abcZ: 404 abcZ: 213 abcZ: 166 abcZ: 114	99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77	1 1 1 1 1 1 1 1 1	0 0 0 0 0 0	433/433 433/433 433/433 433/433 433/433 433/433 433/433	Compare abc2: 657 Compare abc2: 453 Compare abc2: 405 Compare abc2: 404 Compare abc2: 213 Compare abc2: 166 Compare abc2: 114					
Allele abcZ: 657 abcZ: 453 abcZ: 405 abcZ: 404 abcZ: 213 abcZ: 166 abcZ: 114 abcZ: 103	99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77	1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0	433/433 433/433 433/433 433/433 433/433 433/433 433/433 433/433	Compare abc2: 657 Compare abc2: 453 Compare abc2: 453 Compare abc2: 405 Compare abc2: 404 Compare abc2: 213 Compare abc2: 166 Compare abc2: 114 Compare abc2: 103					
Allele abcZ: 657 abcZ: 453 abcZ: 405 abcZ: 404 abcZ: 213 abcZ: 166 abcZ: 114	99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77           99.77	1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0	433/433 433/433 433/433 433/433 433/433 433/433 433/433	Compare abc2: 657 Compare abc2: 453 Compare abc2: 405 Compare abc2: 404 Compare abc2: 213 Compare abc2: 166 Compare abc2: 114					

Click the appropriate 'Compare' button to display a list of nucleotide differences and/or a sequence alignment.



### 11.5.2 Sequence comparison

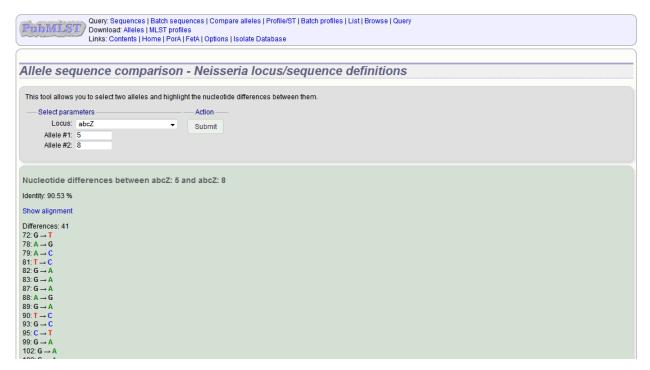
To directly compare two sequences click 'Sequence comparison' from the contents page of a sequence definition database.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database							
Neisseria locus/sequence definitions database							
The Neisseria PubMLST sequence definition database contains allele and profile data representing the this database should have a corresponding record in the isolate database.	he total known diversity of Neisseria species. Every new ST deposited in						
Query database           • Sequence query - query an allele sequence.           • Batch sequence query - query multiple sequences in FASTA format.           • Sequence attribute search - find alleles by matching criteria (all loci together)           • Locus-specific sequence attribute search - select, analyse and download specific alleles.           • Search, browse or enter list of MLST profiles           • Search by combinations of MLST alleles - including partial matching.           • Batch profile query - lookup MLST profiles copied from a spreadsheet.           • Extract finetype from whole genome data	<ul> <li>Downloads         <ul> <li>Allele sequences</li> <li>MLST profiles</li> <li>Set general options</li> <li>Set general options</li> <li>Manage submissions</li> </ul> </li> <li>General information         <ul> <li>Number of sequences: 124770</li> <li>Number of profiles (MLST): 10058</li> <li>Last updated: 2015-08-19</li> <li>Profile update history</li> <li>About BIGSdb</li> </ul> </li> </ul>						
Sequence comparison displ	uences most similar to selected allele. lay a comparison between two sequences. sing allele sequences stored for particular locus.						

Enter the locus and two allele identifiers to compare. Press submit.

Cuery: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database							
Allele sequence comparison - Neisseria locus/sequence definitions							
This tool allows you to select two alleles and highlight the nucleotide differences between them.							
Select parameters     Action       Locus:     abcZ     ▼       Allele #1:     5       Allele #2:     8							

A list of nucleotide differences and/or an alignment will be displayed.



#### See also:

Locus explorer plugin.

## 11.6 Browsing isolate data

Isolate records can be browsed by clicking the link to 'Search or browse database'.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For eve profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database that it does not represent a population sample. Query database Search or browse database Search or browse database Search oy combinations of loci (profiles) Projects - main projects defined in database. Option settings Set display and query options for locus, schemes or scheme fields.	
Breakdown       Export       Export dataset       Codon usage         • Two field       • Contigs       • Contigs       • Codon usage         • Dublications       • Sequences - XMFA / concatenated FASTA formats       • Codon usage         • Publications       • Sequence bin       • Description of database fields	

Leave query form fields blank (the display of these may vary depending on modification options set by the user). Choose the field to order the results by, the number of results per page to display, and click 'Submit'.

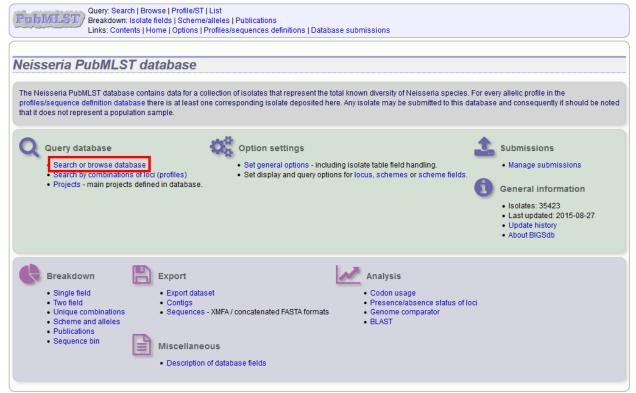
Pu	<b>bMLS</b> 7		s   Scheme/allele		blications uences definitions   Databas	e submissions							
							Help 🗗	Toggle	: 🚯 Field help: id			•	Go
Sea	arch or	browse Neisse	ria PubM	LS	T database								
Ente	er search crite	eria or leave blank to browse	all records.										Modify
	Isolate prove	enance/phenotype fields				ut ontions				- Act	ion		form
	id		Enter value		+ () Order by				- ascending				
	iu .	• - •	Linter value		Display					Re	set Su	bmit	_
					Dispia	25 v records per p	page 🕕						
3534	49 records ret	turned (1 - 25 displayed). Cli	ck the hyperlinks	for de	tailed information.								
				_									
Page	e: 1 2	3 4 5 6 7 8	9 > La	st									
Page	e: 1 2	3 4 5 6 7 8	9 > La:	st									
Page	e: 1 2	3 4 5 6 7 8			elds ()				MLST	Finet	yping antig	ens	
Page id	e: 1 2	3 4 5 6 7 8 aliases				species	serogroup	ST	MLST clonal complex		yping antig PorA VR2		
			Iso	late fi year			serogroup A	ST 4					
id	isolate	aliases	lso country	late fi year 1937	disease	Neisseria meningitidis		_	clonal complex	PorA VR1	PorA VR2	FetA VR	
id 1 2	isolate A4/M1027	aliases B1; NIBSC_2803; Z1001	lso country USA	late fi year 1937 1967	disease invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A A	4	clonal complex ST-4 complex/subgroup IV	PorA VR1 5-2	PorA VR2 10	FetA VR F1-5	
id 1 2	isolate A4/M1027 120M	aliases B1; NIBSC_2803; Z1001	Iso country USA Pakistan	late fi year 1937 1967 2000	disease invasive (unspecified/other) meningitis and septicaemia	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A	4 1	clonal complex ST-4 complex/subgroup IV	PorA VR1 5-2 5-2	PorA VR2 10 10	FetA VR F1-5	
id 1 2 3 4	isolate A4/M1027 120M M00242905	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Iso country USA Pakistan UK	late fi year 1937 1967 2000 1937	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B	4 1 1099 4	cional complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II	PorA VR1 5-2 5-2 19	PorA VR2 10 10	FetA VR F1-5	
id 1 2 3 4 5	isolate A4/M1027 120M M00242905 M1027	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Iso Country USA Pakistan UK USA	late fi year 1937 1967 2000 1937 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A	4 1 1099 4	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup IV	PorA VR1 5-2 5-2 19	PorA VR2 10 10 15	FetA VR F1-5	
id 1 2 3 4 5	isolate A4/M1027 120M M00242905 M1027 M00240227	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Iso country USA Pakistan UK USA UK	late fi year 1937 1967 2000 1937 2000 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B	4 1 1099 4 1100	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup IV ST-32 complex/ET-5 complex	PorA VR1 5-2 5-2 19	PorA VR2 10 10 15	FetA VR F1-5	
id 1 2 3 4 5 6 7	isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207	aliases B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043	Isc country USA Pakistan UK USA UK UK	late fi year 1937 1967 2000 1937 2000 2000 1975	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B W	4 1099 4 1100 1101	cional complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup I/V ST-32 complex/ET-5 complex ST-22 complex	PorA VR1 5-2 5-2 19 7	PorA VR2 10 10 15 16	FetA VR F1-5 F5-1	
id 1 2 3 4 5 6 7	isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891	aliases B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043	Iso country USA Pakistan UK USA UK UK Finland UK	late fi year 1937 2000 1937 2000 2000 1975 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A	4 1099 4 1100 1101 5	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I ST-22 complex/ET-5 complex ST-22 complex ST-5 complex/subgroup III	PorA VR1 5-2 5-2 19 7	PorA VR2 10 10 15 16 9	FetA VR F1-5 F5-1	
id 1 2 3 4 5 6 7 8	isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891 M00242007	aliases B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043	Iso country USA Pakistan UK USA UK UK Finland UK	late fi year 1937 2000 1937 2000 2000 1975 2000 1984	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A B	4 1099 4 1100 1101 5 1102	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup IV ST-32 complex/ET-5 complex ST-22 complex ST-5 complex/subgroup III ST-18 complex	PorA VR1 5-2 5-2 19 7	PorA VR2 10 10 15 16 9	FetA VR F1-5 F5-1	
id 1 2 3 4 5 6 7 8 9	isolate A4/M1027 120M M00242905 M1027 M00240227 M00240227 7891 M00242007 0021/84	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054	Iso country USA Pakistan UK USA UK VK Finland UK Czech Republic	late fi year 1937 2000 1937 2000 2000 1975 2000 1975 2000 1984 1971	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A B W W	4 1099 4 1100 1101 5 1102 114	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IVI ST-4 complex/subgroup IVI ST-32 complex/ET-5 complex ST-22 complex/subgroup III ST-18 complex/subgroup III ST-18 complex ST-22 complex	PorA VR1 5-2 5-2 19 7 20	PorA VR2 10 10 15 16 9 14	FetA VR F1-5 F5-1 F3-1	
id 1 2 3 4 5 6 7 8 9 10	isolate A4/M1027 120M M00242905 M1027 M00240227 M002402207 7891 M00242007 002184 6748	aliases B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043 B54; NIBSC_2760; 21054 B73; NIBSC_2784; 21073	Iso country USA Pakistan UK USA UK Finland UK Czech Republic Canada Germany	late fi year 1937 2000 1937 2000 2000 1975 2000 1975 2000 1984 1971 1964	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis	A B A B W A B W A	4 1099 4 1100 1101 5 1102 114 1 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IV ST-3 complex/subgroup IV ST-32 complex/subgroup IV ST-5 complex/subgroup III ST-16 complex ST-22 complex ST-1 complex/subgroup I/II	PorA VR1 5-2 19 7 20 18-1 5-2	PorA VR2 10 10 15 16 9 14 3	FetA VR F1-5 F5-1 F3-1 F5-1	
id 1 2 3 4 5 6 7 8 9 10 11	isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891 M00242007 0021/84 6748 129E	aliases B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043 B54; NIBSC_2760; 21054 B73; NIBSC_2784; 21073	Iso country USA Pakistan UK USA UK Finland UK Czech Republic Canada Germany	late fi year 1937 2000 1937 2000 2000 1975 2000 1975 2000 1984 1971 1964	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis Neisseria meningiltidis	A B A B W A B W A A A	4 1099 4 1100 1101 5 1102 114 1 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup I/I ST-32 complex/ET-5 complex ST-22 complex/ST-5 Complex/subgroup III ST-18 complex ST-2 complex/subgroup I/II ST-1 complex/subgroup I/II	PorA VR1 5-2 19 7 20 18-1 5-2	PorA VR2 10 10 15 16 9 14 3 10	FetA VR F1-5 F5-1 F3-1 F5-1	
id 1 2 3 4 5 6 7 8 9 10 11 12	isolate A4M1027 120M M00242905 M1027 M00240227 M00240227 7891 M00242007 0021/84 6748 129E 0090/89	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054 B73; NIBSC_2784; Z1073 B92; NIBSC_2828; Z1092	Iso country USA Pakistan UK USA UK Finland UK Czech Republic Canada Germany Czech Republic Philippines	late fi year 1937 2000 1937 2000 2000 1975 2000 1975 2000 1974 1971 1964 1989 1968	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningittäis Neisseria meningittäis	A B A B W A B W A A A B	4 1099 4 1100 1101 5 1102 114 1 1015	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IV ST-4 complex/subgroup IV ST-32 complex/ET-5 complex ST-52 complex/subgroup III ST-18 complex ST-18 complex ST-11 complex/subgroup I/II ST-11 complex/subgroup I/II ST-22 complex/ST-5 complex/ST-22 complex/ST-5 cmplex/ST-5 cmplex/S	PorA VR1 5-2 5-2 19 7 20 18-1 5-2 7	PorA VR2 10 10 15 16 9 14 3 10 16	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6	
id 1 2 3 4 5 6 7 8 9 10 11 12 13	isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891 M00242007 0021/84 6748 129E 0090/89 139M	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054 B73; NIBSC_2784; Z1073 B92; NIBSC_2828; Z1092	Iso country USA Pakistan UK USA UK Finland UK Czech Republic Canada Germany Czech Republic Philippines	late fi year 1937 2000 1937 2000 2000 1975 2000 1975 2000 1974 1971 1964 1989 1968	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningittäis Neisseria meningittäis	A B A B W A B W A A A B A	4 1099 4 1100 1101 5 1102 114 1 1015 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IV ST-4 complex/subgroup IV ST-32 complex/ET-5 complex ST-52 complex/subgroup III ST-18 complex ST-18 complex ST-11 complex/subgroup I/II ST-11 complex/subgroup I/II ST-22 complex/ST-5 complex/ST-22 complex/ST-5 cmplex/ST-5 cmplex/S	PorA VR1 5-2 5-2 19 7 20 18-1 5-2 7	PorA VR2 10 10 15 16 9 14 3 10 16 10	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6	
id 1 2 3 4 5 6 7 8 9 10 11 11 12 13 14	isolate A4/M1027 120M M00242905 M1027 M00240227 M0024027 7891 M002422007 0021/84 129E 0090/89 139M 0120/95	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054 B73; NIBSC_2784; Z1073 B92; NIBSC_2828; Z1092	Iso country USA Pakistan UK USA UK Finland UK Czech Republic Canada Germany Czech Republic Philippines Czech Republic	late fi year 1937 1967 2000 1937 2000 1937 2000 1975 2000 1975 2000 1984 1971 1964 1989 1968 1995	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) carrier	Neisseria meningittäis Neisseria meningittäis	A B B W A B W A A B A A A X	4 1099 4 1100 1101 5 1102 114 1 1015 1 117	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IV ST-4 complex/subgroup IV ST-32 complex/ET-5 complex ST-52 complex/subgroup III ST-18 complex ST-18 complex ST-11 complex/subgroup I/II ST-11 complex/subgroup I/II ST-22 complex/ST-5 complex/ST-22 complex/ST-5 cmplex/ST-5 cmplex/S	PorA VR1 5-2 5-2 19 7 20 18-1 5-2 7	PorA VR2 10 10 15 16 9 14 3 10 16 10	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6	

Clicking the hyperlink for any record will display full information about the profile.

-							Help 🖉	T	: (1) Field help: id			
_							Help (2)	loggle	: 😈 Fiela neip: id			
28	arch or	browse Neisse	eria PubM	LS	T database							
nte	r search crite	ria or leave blank to browse	e all records.									
	Isolate prove	nance/phenotype fields			Display/so	ort options				— — Act	ion	
	d		Enter value		+ 0 Order by							
	u	• - •	Linter value		Display				- ascending	Re	set Su	bmit
					Diopid	, 20 + records per	pago 🕖					
					elds ()				MLST		typing antig	
_	isolate	aliases	country	year	disease	species	serogroup		clonal complex	PorA VR1	PorA VR2	FetA VR
1	A4/M1027	B1; NIBSC_2803; Z1001	country USA	year 1937	disease invasive (unspecified/other)	Neisseria meningitidis	A	4	clonal complex ST-4 complex/subgroup IV	PorA VR1 5-2	PorA VR2 10	FetA VR F1-5
1 2	A4/M1027 120M		country USA Pakistan	year 1937 1967	disease invasive (unspecified/other) meningitis and septicaemia	Neisseria meningitidis Neisseria meningitidis	A A	4 1	clonal complex	PorA VR1 5-2 5-2	PorA VR2 10 10	FetA VR
1 2 3	A4/M1027 120M M00242905	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Country USA Pakistan UK	year 1937 1967 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B	4 1 1099	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II	PorA VR1 5-2	PorA VR2 10	FetA VR F1-5
1 2 3 4	A4/M1027 120M M00242905 M1027	B1; NIBSC_2803; Z1001	Country USA Pakistan UK USA	year 1937 1967 2000 1937	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A	4 1 1099 4	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup IV	PorA VR1 5-2 5-2 19	PorA VR2 10 10 15	FetA VR F1-5
1 2 3 4 5	A4/M1027 120M M00242905 M1027 M00240227	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Country USA Pakistan UK USA UK	year 1937 1967 2000 1937 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B	4 1099 4 1100	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup IV ST-32 complex/ET-5 complex	PorA VR1 5-2 5-2 19	PorA VR2 10 10	FetA VR F1-5
1 2 3 4 5 6	A4/M1027 120M M00242905 M1027 M00240227 M00282207	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043	Country USA Pakistan UK USA UK UK	year 1937 1967 2000 1937 2000 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B W	4 1099 4 1100 1101	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup I/I ST-32 complex/ET-5 complex ST-22 complex/ET-5 complex	PorA VR1 5-2 5-2 19 7	PorA VR2 10 10 15 16	FetA VR F1-5 F5-1
1 2 3 4 5 6 7	A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Country USA Pakistan UK USA UK UK Finland	year 1937 1967 2000 1937 2000 2000 1975	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A	4 1099 4 1100 1101 5	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-4 complex/subgroup IV ST-32 complex/ET-5 complex ST-22 complex ST-5 complex/subgroup III	PorA VR1 5-2 5-2 19	PorA VR2 10 10 15 16 9	FetA VR F1-5
1 2 3 4 5 6 7 8	A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891 M00242007	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043	Country USA Pakistan UK USA UK UK Finland UK	year 1937 1967 2000 1937 2000 2000 1975 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A B	4 1099 4 1100 1101 5 1102	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I ST-32 complex/ET-5 complex ST-52 complex/subgroup III ST-18 complex	PorA VR1 5-2 5-2 19 7	PorA VR2 10 10 15 16	FetA VR F1-5 F5-1
1 2 3 4 5 6 7 8 9	A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891 M00242007 0021/84	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054	Country USA Pakistan UK USA UK Finland UK Czech Republic	year 1937 2000 1937 2000 2000 1975 2000 1975 2000 1984	disease invasive (unspecified/other) meningitis and septicarenia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A B W W	4 1099 4 1100 1101 5 1102 114	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IV ST-4 complex/subgroup IV ST-32 complex/ET-5 complex ST-22 complex/ST-5 complex/subgroup III ST-18 complex ST-22 complex ST-22 complex	PorA VR1 5-2 5-2 19 7 20	PorA VR2 10 10 15 16 9 14	FetA VR F1-5 F5-1 F3-1
2 3 4 5 6 7 8 9 10	A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891 M00242007 0021/84 6748	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054 B73; NIBSC_2784; Z1073	Country USA Pakistan UK USA UK VK Finland UK Czech Republic Canada	year 1937 1967 2000 1937 2000 2000 1975 2000 1984 1971	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A B W A	4 1099 4 1100 1101 5 1102 114 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IVI ST-4 complex/subgroup IVI ST-32 complex/ET-5 complex ST-52 complex ST-5 complex/subgroup III ST-18 complex ST-22 complex ST-12 complex	PorA VR1 5-2 5-2 19 7 20 18-1	PorA VR2 10 10 15 16 9 14 3	FetA VR F1-5 F5-1 F3-1 F5-1
1 2 3 4 5 6 7 8 9 10 11	A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891 M00242007 0021/84 6748 129E	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054	country USA Pakistan UK USA UK Finland UK Czech Republic Canada Germany	year 1937 1967 2000 1937 2000 2000 1975 2000 1984 1971 1964	disease invasive (unspecified/other) meningitis and septicaremia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis	A B A B W A B W A A A	4 1099 4 1100 1101 5 1102 114 1 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I ST-22 complex/ET-5 complex ST-52 complex/subgroup III ST-18 complex ST-22 complex ST-22 complex ST-12 complex/subgroup I/I ST-1 complex/subgroup I/I	PorA VR1 5-2 19 7 20 18-1 5-2	PorA VR2 10 10 15 16 9 14 3 10	FetA VR F1-5 F5-1 F3-1
1 2 3 4 5 6 7 8 9 10 11 12	A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891 M00242007 0021/84 6748 129E 0090/89	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054 B73; NIBSC_2764; Z1073 B92; NIBSC_2828; Z1092	Country USA Pakistan UK USA UK Finland UK Czech Republic Germany Czech Republic	year 1937 1967 2000 1937 2000 2000 1975 2000 1975 2000 1984 1971 1964 1989	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis Neisseria meningittidis	A B A B W A B W A A A B	4 1099 4 1100 1101 5 1102 114 1 1 1015	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IV ST-3 complex/subgroup IV ST-32 complex/ET-5 complex ST-22 complex/subgroup III ST-18 complex/subgroup I/II ST-1 complex/subgroup I/II ST-1 complex/subgroup I/II ST-32 complex/ET-5 complex	PorA VR1 5-2 5-2 19 7 20 18-1 5-2 7	PorA VR2 10 10 15 16 9 14 3 10 16	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6
1 2 3 4 5 6 7 8 9 10 11 12 13	A4/M1027 120M M00242905 M1027 M00240227 7891 M00282207 7891 M00242007 0021/84 6748 129E 0090/89 139M	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054 B73; NIBSC_2784; Z1073	Country USA Pakistan UK USA UK Finland UK Czech Republic Canada Germany Czech Republic Philippines	year 1937 1967 2000 1937 2000 2000 1975 2000 1975 2000 1984 1971 1964 1989 1968	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningittidis Neisseria meningittidis	A B A B W A B W A A A B A A	4 1099 4 1100 1101 5 1102 114 1 1 1015 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I ST-22 complex/ET-5 complex ST-52 complex/subgroup III ST-18 complex ST-22 complex ST-22 complex ST-12 complex/subgroup I/I ST-1 complex/subgroup I/I	PorA VR1 5-2 19 7 20 18-1 5-2	PorA VR2 10 10 15 16 9 14 3 10 16 10	FetA VR F1-5 F5-1 F3-1 F5-1
1 2 3 4 5 6 7 7 8 9 10 11 12 13 14	A4/M1027 120M M00242905 M1027 M00240227 7891 M00282207 7891 M00242007 0021/84 6748 129E 0090/89 139M 0120/95	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054 B73; NIBSC_2764; Z1073 B92; NIBSC_2828; Z1092	Country USA Pakistan UK UK VK Finland UK Czech Republic Canada Germany Czech Republic Philippines Czech Republic	year 1937 1967 2000 1937 2000 2000 1975 2000 1975 2000 1984 1971 1964 1989 1968 1995	disease invasive (unspecified/other) meningitis and septicaremia invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningittidis Neisseria meningittidis	A B B W A B W A B W A A A A X	4 1099 4 1100 1101 5 1102 114 1 1 1015 1 117	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IVI ST-4 complex/subgroup IVI ST-32 complex/ET-5 complex ST-22 complex/subgroup III ST-18 complex/subgroup IVI ST-1 complex/subgroup IVI ST-1 complex/subgroup IVI ST-32 complex/ET-5 complex	PorA VR1 5-2 5-2 19 7 20 18-1 5-2 7	PorA VR2 10 10 15 16 9 14 3 10 16	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6
1 2 3 4 5 6 7 8 9 10 11 12 13	A4/M1027 120M M00242905 M1027 M00240227 7891 M00282207 7891 M00242007 0021/84 6748 129E 0090/89 139M	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760; Z1054 B73; NIBSC_2764; Z1073 B92; NIBSC_2828; Z1092	Country USA Pakistan UK UK VK Finland UK Czech Republic Canada Germany Czech Republic Philippines Czech Republic	year 1937 1967 2000 1937 2000 2000 1975 2000 1975 2000 1984 1971 1964 1989 1968	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningittidis Neisseria meningittidis	A B B W A B W A A B A A A E	4 1099 4 1100 1101 5 1102 114 1 1 1015 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup IVI ST-4 complex/subgroup IVI ST-32 complex/ET-5 complex ST-22 complex/subgroup III ST-18 complex/subgroup IVI ST-1 complex/subgroup IVI ST-1 complex/subgroup IVI ST-32 complex/ET-5 complex	PorA VR1 5-2 5-2 19 7 20 18-1 5-2 7	PorA VR2 10 10 15 16 9 14 3 10 16 10	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6

# 11.7 Querying isolate data

The 'Search or browse database' page of an isolate database allows you to also search by combinations of provenance criteria, scheme and locus data, and more.



To start with, only one provenance field search box is displayed but more can be added by clicking the '+' button (highlighted). These can be linked together by 'and' or 'or'.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Pul Links: Contents   Home   Options   Profiles/sequ			
		Toggle: <i>i</i> Field help: id	▼ Go
Search Neisseria PubMLST database			
Isolate provenance/phenotype fields Combine with: AND • country • = • USA year • > • 1999 Action Reset Submit	Display/sort options Order by: id	✓ ascending ✓	Modify form options

After the search has been submitted, the results will be displayed in a table.

Pul	MILST) Break	down: Isola	te fielo	Profile/ST   List Is   Scheme/alleles   Publicat   Options   Profiles/sequence		submissio	าร						
								Toggle: i	ield help:	id		<b>-</b>	Go
Seal	rch Neisser	ia Pub	ML	ST database									
—-Is	olate provenance/ph	enotype fiel	ds —		Display/s	ort options –						Mo	dify
Con	nbine with: AND 👻				Order b	y: id			ending	•		form	
CO	ountry 🗸			- USA	+ i Displa	y: 25 👻	record	Is per page i				opt	
ye	ar 🗸	- >		1999								_	
— A	ction												
R	eset Submit												
300 re	cords returned (1 - 25	displayed)	. Click	the hyperlinks for detailed inf	ormation.								
_													
Page:	1234	5 6	78	9 > Last									
				lsolate fields 👔				MLST	Fine	typing antig	ens		
id	isolate alias	es country	year	disease	species	serogroup	ST	clonal complex	PorA VR1	PorA VR2	FetA VR		
341	M7085	USA		invasive (unspecified/other)		W	11						
499	MDL01A0601	USA	2001	meningitis	Neisseria meningitidis	Y	1378	ST-23 complex/Cluster A3					
500	MDL01A2447	USA	2001	invasive (unspecified/other)	Neisseria meningitidis	Y	1379	ST-23 complex/Cluster A3					
866	MD01227	USA	2001		Neisseria meningitidis		1624	ST-167 complex					
867	MDO1056	USA	2001		Neisseria meningitidis		1625	ST-23 complex/Cluster A3					
868	MDO1066	USA	2001		Neisseria meningitidis		1626	ST-269 complex					
2281	M7089	USA	2000	invasive (unspecified/other)	Neisseria meningitidis	W	11	ST-11 complex/ET-37 complex	5	2			
2299	M7257	USA	2000	invasive (unspecified/other)	Neisseria meningitidis	W	11	ST-11 complex/ET-37 complex	5	2			
2316	M7086	USA	2000	invasive (unspecified/other)	Neisseria meningitidis	W	22	ST-22 complex	6	3			
2317	M7084	USA	2000	invasive (unspecified/other)	Neisseria meningitidis	W	22	ST-22 complex	6	3			
2322	M7092	USA	2000	invasive (unspecified/other)	Neisseria meningitidis	W	1065	ST-22 complex	6	3			
2323	M7100	USA	2000	invasive (unspecified/other)	Neisseria meningitidis	W	1286	ST-22 complex	6	3			
2324	M7259	USA	2000	invasive (unspecified/other)	Neisseria meningitidis	W	1066	ST-22 complex	6	3			

Each field can be queried using standard operators.

More search features are available by clicking the 'Modify form options' tab on the right-hand side of the screen.

Lisolate provenance/phenotype fields       Display/sort options         Combine with: AND ↓       Order by: id       ascending ↓         Country ↓ = ↓ USA       ↓ I       Display: 25 ↓ records per page []         year ↓ > ↓ 1999       Action       Reset       Submit         0 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.       ascending ↓       Last         1       Isolate aliases country year       disease       species       serogroup         1       Isolate fields []       MLST       Finetyping antigens         14       Kolon       Noiseria meningitidis       W       11       ST-11 complex       PortA VR1 Port VR2 FetA VR         14       Kolon       USA 2000 invasive (unspecified/other) Neisseria meningitidis       W       11       ST-11 complex/E-137 complex         14       M7085       USA 2000 invasive (unspecified/other) Neisseria meningitidis       Y       1378       ST-23 complex/Cluster A3         199       MDL01A0601       USA 2001 invasive (unspecified/other) Neisseria meningitidis       Y       1378       ST-23 complex/Cluster A3         199       MDL01A0601       USA 2001 invasive (unspecified/other) Neisseria meningitidis       Y       1378       ST-23 complex/Cluster A3
Combine with: AND Combine with: AND Country year Action Reset Submit Display: 25 records per page Action Reset Submit Display: 25 records per page 25 records per page 27 records per page 27
Combine with: AND  Country Cou
country       =       USA       + I       Display: 25 • records per page []         year       >       >       1999         Action
year v > v 1999 Action Reset Submit D records returned (1 - 25 displayed). Click the hyperlinks for detailed information. ge: 1 2 3 4 5 6 7 8 9 > Last Isolate fields I MLST Finetyping antigens Isolate fields I Second Second Part Park VR1 Park VR2 FetA VR 141 M7085 USA 2000 invasive (unspecified/other). Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 199 MDL01A0601 USA 2000 invasive (unspecified/other). Neisseria meningitidis Y 1378 ST-23 complex/CLUSEr A3
Action Reset Submit Precords returned (1 - 25 displayed). Click the hyperlinks for detailed information. ge: 1 2 3 4 5 6 7 8 9 > Lest Isolate fields 1 KIN Finetyping antigens Isolate fields 2 Species Serogroup Complex PorA VR1 PorA VR2 FetA VR 41 M7085 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 9 MDL01A0601 USA 2001 meningitis Neisseria meningitidis Y 1378 ST-23 complex/Cluster A3
Reset Submit Precords returned (1 - 25 displayed). Click the hyperlinks for detailed information. ge: 1 2 3 4 5 6 7 8 9 > Last Isolate fields : MLST Finetyping antigens Isolate fields : Secontry year disease species sergroup ST clonal complex PorA VR1 PorA VR2 [FetA VR VR2 [FetA VR 99 MDL01A0601 USA 2001 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 99 MDL01A0601 USA 2001 meningitis Neisseria meningitidis Y 1378 ST-23 complex/Cluster A3
D records returned (1 - 25 displayed). Click the hyperlinks for detailed information. ge: 1 2 3 4 5 6 7 8 9 > Lest Isolate fields 2 MLST Finetyping antigens Isolate aliases country year disease species serogroup ST clonal complex PorA VR1 PorA VR2 FetA VR 141 M7085 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 199 MDL01A0601 USA 2001 meningitis Neisseria meningitidis Y 1378 ST-23 complex/Cluster A3
D records returned (1 - 25 displayed). Click the hyperlinks for detailed information. ge: 1 2 3 4 5 6 7 8 9 > Lest Isolate fields 2 MLST Finetyping antigens id isolate aliases country year disease species serogroup ST clonal complex PorA VR1 PorA VR2 FetA VR H1 M7085 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 199 MDL01A0601 USA 2001 meningitis Neisseria meningitidis Y 1378 ST-23 complex/Cluster A3
ge: 1 2 3 4 5 6 7 8 9 > Lest Isolate fields 2 MLST Finetyping antigens id isolate aliases country year disease species serogroup ST clonal complex PorA VR1 PorA VR2 FetA VR INTOR5 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 99 MDL01A0601 USA 2001 meningitis Neisseria meningitidis Y 1378 ST-23 complex/Cluster A3
ge: 1 2 3 4 5 6 7 8 9 > Lest Isolate fields 2 MLST Finetyping antigens id isolate aliases country year disease species serogroup ST clonal complex PorA VR1 PorA VR2 FetA VR 41 M7085 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 99 MDL01A0601 USA 2001 meningitis Neisseria meningitidis Y 1378 ST-23 complex/Cluster A3
ge: 1 2 3 4 5 6 7 8 9 > Lest Isolate fields 2 MLST Finetyping antigens id isolate aliases country year disease species serogroup ST clonal complex PorA VR1 PorA VR2 FetA VR INTOR5 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 99 MDL01A0601 USA 2001 meningitis Neisseria meningitidis Y 1378 ST-23 complex/Cluster A3
Isolate fields II         MLST         Finetyping antigens           isolate         aliases         country         year         disease         serogroup         ST         clonal complex         PorA VR1         PorA VR2         FetA VR           141         M7085         USA         2000         invasive (unspecified/other)         Neisseria meningitidis         W         11         ST-11 complex/ET-37 complex           199         MDL0140601         USA         2001         meningitis         Neisseria meningitidis         Y         1378         ST-23 complex/Cluster A3
Isolate fields I         MLST         Finetyping antigens           isolate         aliases country year         disease         species         serogroup         ST         clonal complex         PorA VR1         PorA VR2         FetA VR           141         M7085         USA         2000         invasive (unspecified/other)         Neisseria meningitidis         W         11         ST-11 complex/ET-37 complex           199         MDL01A0601         USA         2001         meningitidis         Y         1378         ST-23 complex/Cluster A3
id         isolate         aliases         country         year         disease         species         serogroup         ST         clonal complex         PorA VR1         PorA VR2         FetA VR           441         M7085         USA         2000         invasive (unspecified/other)         Neisseria meningitidis         W         11         ST-11 complex/ET-37 complex           99         MDL01A0601         USA         2001         meningitidis         Neisseria meningitidis         Y         1378         ST-23 complex/Cluster A3
id         isolate         aliases         country         year         disease         species         serogroup         ST         clonal complex         PorA VR1         PorA VR2         FetA VR           V41         M7085         USA         2000         invasive (unspecified/other)         Neisseria meningitidis         W         11         ST-11 complex/ET-37 complex         FetA VR2         FetA VR2         FetA VR2         FetA VR2         FetA VR2         FetA VR2         FetA VR3
41         M7085         USA         2000         invasive (unspecified/other)         Neisseria meningitidis         W         11         ST-11 complex/ET-37 complex           99         MDL0140601         USA         2001         meningitis         Neisseria meningitidis         Y         1378         ST-23 complex/Cluster A3
99 MDL01A0601 USA 2001 meningitis Neisseria meningitidis Y 1378 ST-23 complex/Cluster Å3
100 MDL 0142447 LISA 2001 invasive (upspecified/other) Neisseria meninaitidis V 1270 ST 22 complex/Cluster 42
Wide Mode and a state (unspecifie double) Nelssena meningidals 1 1575 31-25 complex claster AS
66 MDO1227 USA 2001 Neisseria meningitidis 1624 ST-167 complex
67 MDO1056 USA 2001 Neisseria meningitidis 1625 ST-23 complex/Cluster A3
168 MD01066 USA 2001 Neisseria meningitidis 1626 ST-269 complex
281 M7089 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 5 2
299 M7257 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex/ET-37 complex 5 2
316 M7086 USA 2000 invasive (unspecified/other) Neisseria meningitidis W 22 ST-22 complex 6 3
316         M7086         USA         2000 invasive (unspecified/other)         Neisseria meningitidis         W         22         ST-22 complex         6         3           317         M7084         USA         2000 invasive (unspecified/other)         Neisseria meningitidis         W         22         ST-22 complex         6         3
316         M7086         USA         2000 invasive (unspecified/other)         Neisseria meningitidis         W         22         ST-22 complex         6         3

A tab will be displayed. Different options will be available here depending on the database. Queries will be combined from the values entered in all form sections. Possible options are:

- · Provenance fields
  - Search by combination of provenance field values, e.g. country, year, sender.
- Allele designations/scheme field values
  - Search by combination of allele designations and/or scheme fields e.g. ST, clonal complex information.
- Allele designation status
  - Search by whether allele designation status is confirmed or provisional.
- Tagged sequence status
  - Search by whether tagged sequence data is available for a locus. You can also search by sequence flags.
- Attribute values list
  - Enter a list of values for any provenance field, locus, or scheme field.
- Filters
  - Various filters may be available, including
    - \* Publications
    - \* Projects
    - \* MLST profile completion status
    - \* Clonal complex
    - \* Sequence bin size
    - \* Inclusion/exclusion of old versions

PribNILST         Query: Search   Profile/ST           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions								
	Field help: id							
Search or browse Neisseria PubMLST database								
Enter search criteria or leave blank to browse all records. Modify form parameters to fil — Isolate provenance/phenotype fields id	ter Modify Modify form parameters options							
Display/sort options	Click to add or remove additional query terms:							
Order by: id vascending Display: 25 v records per page 1	<ul> <li>Provenance fields</li> <li>Allele designations/scheme field values</li> <li>Allele designation counts</li> <li>Allele designation status</li> <li>Tagged sequence counts</li> <li>Tagged sequence status</li> <li>Attribute values list</li> <li>Filters</li> </ul>							

If the interface is modified, a button to save options becomes available within the tab. If this is clicked, the modified form will be displayed the next time you go to the query page.

## 11.7.1 Query by allele designation/scheme field

Queries can be combined with allele designation/scheme field values.

Make sure that the allele designation/scheme field values fieldset is displayed by selecting it in the 'Modify form options' tab.

Profit MILST Query: Search   Profile/ST Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
Help [27] Toggle: (1)	Field help: id 🗸 🗸 🕞
Search or browse Neisseria PubMLST database	
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.  Isolate provenance/phenotype fields id	
Displaysont options Action Order by: id • ascending • Reset Submit Display: 25 • records per page •	Click to add or remove additional query terms:

Designations can be queried using standard operators.

Additional search terms can be combined using the '+' button.

Add your search terms and click 'Submit'. Allele designation/scheme field queries will be combined with terms

TTIPI		ery: Search			le/ST   List heme/alleles   Pub	lications										
~~~~	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,				ons   Profiles/sequ		ns   Datab	ase submi	issio	is						
										Help 🖉	Toggle: 🜖 🛛 Fie	ld help: id				•
ear	ch or bro	wse N	leis	seria	PubMLST	databa	se									
inter se	earch criteria or I	leave blank	to bro	wse all re	cords.											Мо
-Iso	late provenance/	/phenotype	fields				—Allele o	lesionation	s/sch	eme fields						for
	ine with: AND						ST (ML			=	<b>→</b> 11		+	0		ор
cou	ntry	<b>•</b> =		+ USA	4	+ 0	Dianla	de art antia							Action	
year	r	<b>~</b> >		<b>v</b> 199	9			/sort option	ns —						-Action-	
							Orde	r by: id					<ul> <li>ascend</li> </ul>	ding 👻	Reset	Submit
							Dis	play: 25		cords per p	age 🕕					
) recor	ds returned. Clic	k the hyper	links f				Dis	olay: 25	▼ 16							
			_	Isolat	e fields 🜖					M	LST		yping anti <u>c</u>			
id	isolate aliase	es country	year	Isolat	e fields () disease	specie	15	serogroup	ST	M	LST al complex	Finet; PorA VR1				
id 341	isolate aliase M7085	es country USA	year 2000	Isolat invasive (	e fields () disease unspecified/other)	Neisseria me	es ningitidis	serogroup W	<b>ST</b>	M clon ST-11 comp	LST al complex lex/ET-37 complex	PorA VR1	PorA VR2	PetA VR		
id 341 2281	isolate aliase M7085 M7089	es country USA USA	<mark>year</mark> 2000 2000	Isolat invasive ( invasive (	e fields () disease unspecified/other) unspecified/other)	Neisseria me Neisseria me	es ningitidis ningitidis	serogroup W W	<b>ST</b> 11 11	M clon ST-11 comp ST-11 comp	LST al complex lex/ET-37 complex lex/ET-37 complex	PorA VR1	PorA VR2			
id 341 2281 2299	isolate aliase M7085	es country USA USA USA	year 2000 2000 2000	Isolat invasive ( invasive ( invasive (	e fields () disease unspecified/other) unspecified/other) unspecified/other)	Neisseria me Neisseria me Neisseria me	s ningitidis ningitidis ningitidis	serogroup W	ST 11 11 11	M clon. ST-11 comp ST-11 comp ST-11 comp	LST al complex lex/ET-37 complex lex/ET-37 complex lex/ET-37 complex	PorA VR1	PorA VR2	PetA VR		
id 341 2281 2299 19369	isolate aliase M7085 M7089 M7257	es country USA USA USA USA	year 2000 2000 2000 2000	Isolat invasive ( invasive ( invasive ( invasive (	e fields () disease unspecified/other) unspecified/other)	Neisseria me Neisseria me Neisseria me Neisseria me	s ningitidis ningitidis ningitidis ningitidis	serogroup W W W	ST 11 11 11 11	M Clon ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp	LST al complex lex/ET-37 complex lex/ET-37 complex	PorA VR1 5 5 5-1	PorA VR2	FetA VR F1-1		
id 341 2281 2299 19369 19371	isolate aliase M7085 M7089 M7257 M13519	es country USA USA USA USA USA	year 2000 2000 2000 2005 2005	Isolat invasive ( invasive ( invasive ( invasive ( invasive (	e fields () disease unspecified/other) unspecified/other) unspecified/other) unspecified/other)	Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me	ningitidis ningitidis ningitidis ningitidis ningitidis	serogroup W W W	<b>ST</b> 11 11 11 11 11	M Clon ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp	LST al complex lex/ET-37 complex lex/ET-37 complex lex/ET-37 complex lex/ET-37 complex	PorA VR1 5 5 5-1 5-1	PorA VR2 2 2 10-8	FetA VR F1-1		
id 341 2281 2299 19369 19371 19374	isolate aliase M7085 M7089 M7257 M13519 M15141	es country USA USA USA USA USA USA	year 2000 2000 2000 2005 2006 2007	Isolat invasive ( invasive ( invasive ( invasive ( invasive ( invasive (	e fields () disease unspecified/other) unspecified/other) unspecified/other) unspecified/other)	Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me	is ningitidis ningitidis ningitidis ningitidis ningitidis ningitidis	serogroup W W W W C	ST 11 11 11 11 11 11	M Clonn ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp	LST al complex lex/ET-37 complex lex/ET-37 complex lex/ET-37 complex lex/ET-37 complex	PorA VR1 5 5-1 5-1 5-1 5-1	PorA VR2 2 10-8 10-8	FetA VR F1-1		
id 341 2281 2299 19369 19371 19374 19377	isolate aliase M7085 M7089 M7257 M13519 M15141 M16917	es country USA USA USA USA USA USA	year 2000 2000 2000 2005 2006 2007 2008	Isolat invasive ( invasive ( invasive ( invasive ( invasive ( invasive (	e fields () disease unspecified/other) unspecified/other) unspecified/other) unspecified/other) unspecified/other)	Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me	is ningitidis ningitidis ningitidis ningitidis ningitidis ningitidis	serogroup W W W C NG	<b>ST</b> 11 11 11 11 11 11 11	M clon: ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp	LST al complex lewET-37 complex lewET-37 complex lewET-37 complex lewET-37 complex lewET-37 complex	PorA VR1 5 5-1 5-1 5-1 5-1 5-1	PorA VR2 2 10-8 10-8 10-8	FetA VR F1-1		
id 341 2281 2299 19369 19371 19374 19377 19379	isolate aliase M7085 M7089 M7257 M13519 M15141 M16917 M17661	ES COUNTRY USA USA USA USA USA USA USA	year 2000 2000 2005 2005 2006 2007 2008 2009	Isolat invasive ( invasive ( invasive ( invasive ( invasive ( invasive ( invasive (	e fields () disease unspecified/other) unspecified/other) unspecified/other) unspecified/other) unspecified/other) unspecified/other)	Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me Neisseria me	is ningitidis ningitidis ningitidis ningitidis ningitidis ningitidis ningitidis	serogroup W W W W C NG W	ST 11 11 11 11 11 11 11 11	M clon ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp ST-11 comp	LST al complex lewET-37 complex lewET-37 complex lewET-37 complex lewET-37 complex lewET-37 complex lewET-37 complex	PorA VR1 5 5-1 5-1 5-1 5-1 5-1 5-1 5-1 5	PorA VR2 2 10-8 10-8 10-8 10-8 10-8	FetA VR F1-1 F3-6		

### entered in other sections.

### 11.7.2 Query by allele designation count

Queries can be combined with counts of the total number of designations or for individual loci.

Make sure that the allele designation counts fieldset is displayed by selecting it in the 'Modify form options' tab.

Query: Search   Profile/ST Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions Help 12 Toggle: 1 Field	eld help: id 🔹 🐨								
Search or browse Neisseria PubMLST database									
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.  Isolate provenance/phenotype fields Id  Filter value  Display/sort options Order by: id Display: 25  records per page Action Reset Submit	<ul> <li>★ Modify form parameters</li> <li>Click to add or remove additional query terms:</li> <li>● Provenance fields</li> <li>● Allele designations/scheme field values</li> <li>● Allele designation counts</li> <li>● Allele designation status</li> <li>● Allele designation status</li> <li>● Tagged sequence status</li> <li>● Filters</li> </ul>								

For example, to find all isolates that have designations at >1000 loci, select 'total designations > 1000', then click 'Submit'.

Pu	<b>bMLS</b>		ate fields   So		illeles   Publications rofiles/sequences defir	itions								
							Help 🗗	Toggle: 🚯 🛛 Fi	ield help:	id				▼ Go
Sea	rch or	r browse Ne	eisseria	Puk	MLST datab	ase								
Ente	r search cri	teria or leave blank to	browse all re	ecords.	Modify form parameters	s to filter or enter	a list of value	S.						Modify
	solate prov	enance/phenotype fie	elds			- Allele desi	onation coun	ts						form
ic	ł	▼ =		ter valu	e + (	Count of to	tal designatio	ons 👻 >	▼ 100	)	🚽 🛨 🟮			options
	Display/sor	t options				A	ction —							
	Order by:	id				ling 👻 🕞	eset Sub	mit						
	Display:	25 - records pe	r page 🙃											
7473	records ret	urned (1 - 25 displaye	ed). Click the	hyperlin	ks for detailed informat	tion.								
Page	: 1 2	3 4 5 6	7 8 9	) M	Last									
_														
			lso	late fie	lds 🕦			Coghin oizo			MLST	Finety	ping ant	tigens
id	isolate	aliases	country	year	disease	species	serogroup	Seqbin size (bp)	Contigs	ST	clonal complex	PorA VR1	PorA VR2	FetA VR
1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	2069108	364	4	ST-4 complex/subgroup IV	5-2	10	F1-5
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicaemia	Neisseria meningitidis	A	2059411	359	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitidis	A	2057385	199	5	ST-5 complex/subgroup III	20	9	F3-1
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	4241338	652	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
11	129E	B92; NIBSC_2828; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	A	2072690	272	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
12	12014	D00: NIDSC 0705:	Dhilippipoo	4060	(	Neiceorie	٨	01/17/1	202	4	OT 4	E 0	10	CE 4

You can also search for isolates where any isolate has a particular number of designations. Use the term 'any locus' to do this.

Finally, you can search for isolates with a specific number of designations at a specific locus.

Query: Search   Profile/ST Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
Help 🗗 Toggle: (3) Field help: id	Go
Search or browse Neisseria PubMLST database	
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.  — Isolate provenance/phenotype fields —	Modify form
Isolate provenance/phenotype fields     Allele designation counts       id     ▼     =     ▼ Enter value       td     ▼     >       1     ↓     ↑       0     Count of abcZ     ▼	options
Display/sort options     Action       Order by:     id       Display:     25 → records per page ①       1 record returned. Click the hyperlink for detailed information.	
Isolate fields ()     MLST     Finetyping antigens       id     isolate     aliases     country     year     disease     species     serogroup     St     clonal complex     PorA VR1     PorA VR2     FetA VR       34733     LNP15075     Burkina Faso     1997     Neisseria meningitidis     3316868     1010     20     10-1     F3-1	
Analysis tools: Breakdown: Fields Two Field Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status Analysis: Codons Presence/Absence Genome Comparator BLAST Export Dataset Contigs Sequences	

Additional search terms can be combined using the '+' button. Designation count queries will be combined with terms entered in other sections.

**Note:** Searches for 'all loci' with counts that include zero, e.g. 'count of any locus = 0' or with a '<' operator are not supported. This is because such searches have to identify every isolate for which one or more loci are missing. In databases with thousands of loci this can be a very expensive database query.

### 11.7.3 Query by allele designation status

Allele designations can be queried based on their status, i.e. whether they are confirmed or provisional. Queries will be combined from the values entered in all form sections.

Make sure that the allele designation staus fieldset is displayed by selecting it in the 'Modify form options' tab.

Cuery: Search   Profile/ST Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
	Help 🗹 Toggle: 🕄 Field help: id 🗸 🗸 Go
Search or browse Neisseria PubMLST database	
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of Isolate provenance/phenotype fields — Display/sort option id v = v Enter value + O Order by: id	× Modity
	<ul> <li>records per page </li> <li>Click to add or remove additional query terms:</li> <li>Provenance fields</li> <li>Allele designations/scheme field values</li> <li>Allele designation counts</li> <li>Allele designation status</li> <li>Tagged sequence counts</li> <li>Tagged sequence status</li> <li>Attribute values list</li> <li>Filters</li> </ul>

Select a locus from the dropdown box and either 'provisional' or 'confirmed'. Additional query fields can be displayed by clicking the '+' button. Click 'Submit'.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Help 🗹 🔰 Toggle: 🚯 Field help: id	▼ Go
Search or browse Neisseria PubMLST database	
Enter search criteria or leave blank to browse all records.	Modify
	form options
id • = • Entervalue + 0 adk • Is provisional • + 0	opuons
- Display/sort options - Action - Action -	
Order by: id vascending v Reset Submit	
Display: 25 - records per page ()	
1 record returned. Click the hyperlink for detailed information.	
Isolate fields <b>o</b> MLST Finetyping antigens Loci	
id isolate aliases country vear disease species serogroup ST clonal complex PorA VR1 PorA VR2 FetA VR test9	
19363         961-5945         Unknown         Neisseria meningitidis         153         ST-8 complex/Cluster A4         21         16         F5-8	
Analysis tools:	
Breakdown: Fields Two Field Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status	
Analysis: Codons Presence/Absence Genome Comparator BLAST	
Export Dataset Contigs Sequences	

Provisional allele designations are marked within the results tables with a pink background. Any scheme field designations that depend on the allele in question, e.g. a MLST ST, will also be marked as provisional.

## 11.7.4 Query by sequence tag count

Queries can be combined with counts of the total number of tags or for individual loci.

Make sure that the tagged sequence counts fieldset is displayed by selecting it in the 'Modify form options' tab.

Help 🗹 Toggle: 🚯 F	ield help: id 🔹 🖌 Go								
Search or browse Neisseria PubMLST database									
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.  Isolate provenance/phenotype fields Id  I =  Enter value  Order by: Id Display: 25 records per p Action Reset Submit	<ul> <li>Modify form parameters</li> <li>Click to add or remove additional query terms:</li> <li>Provenance fields</li> <li>Allele designations/scheme field values</li> <li>Allele designation counts</li> <li>Allele designation status</li> <li>Allele designation status</li> <li>Tagged sequence counts</li> <li>Tagged sequence status</li> <li>Attribute values list</li> <li>Filters</li> </ul>								

For example, to find all isolates that have sequence tags at >1000 loci, select 'total tags > 1000', then click 'Submit'.

Pul	<b>bMLS</b>		late fields   S		/alleles   Publications Profiles/sequences de	finitions								
						H	lelp 🗹 🛛 To	ggle: 🚺 🛛 Fie	ld help: id					- Go
Sea	rch o	r browse Ne	eisseria	a Pu	bMLST data	base								
Ente	r search cri	iteria or leave blank tr	hrowse all	records	. Modify form paramete	ers to filter or ent	er a list of val	1165						
		/enance/phenotype fie		recorda	. moony form paramete		sequence co							Modify form
ic		venance/prienotype in		nter val	ue +		total tags		> • 1	000	≑ + 8			options
			•	iicei vai	·····	Jount of	-	•	~ 1	000				
	Display/sor						Action							
	Order by:					nding 👻	Reset S	ubmit						
	Display:	25 - records pe	r page 🕦											
7480	records ret	urned (1 - 25 display	ad) Click the	hyperli	inks for detailed inform	ation								
		amea (1 25 alspia)	ou). Onex are	, nypen		auon.								
Page	: 1 2	3 4 5 6	7 8 9		Last									
_														
			lso	plate fie	lds ()			Seabin			MLST	-	ping an	
id	isolate	aliases	country	year	disease	species	serogroup	size (bp)	Contigs	ST	clonal complex	PorA VR1	PorA VR2	FetA VR
1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	2069108	364	4	ST-4 complex/subgroup IV	5-2	10	F1-5
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicaemia	Neisseria meningitidis	A	2059411	359	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitidis	A	2057385	199	5	ST-5 complex/subgroup III	20	9	F3-1
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	4241338	652	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
11	129E	B92; NIBSC_2828; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	A	2072690	272	1	ST-1 complex/subgroup I/II	5-2	10	F3-6

You can also search for isolates where any isolate has a particular number of sequence tags. Use the term 'any locus' to do this.

Finally, you can search for isolates with a specific number of tags at a specific locus.

Pul	Query: Search   Profile/ST           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions													
	Help 🗹 Toggle: 🚯 Field help: id 🗸 Go													
Sea	earch or browse Neisseria PubMLST database													
Enter	Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.													
l												form		
id		<b>▼</b> =	▼ Er	nter va	ue +	<ol> <li>Count of</li> </ol>	BACT00006	5 (rpmJ) 🖣		2	🗧 🛨 📵			options
	Display/sort	options					Action			_				
	Order by:	id			✓ asce	ending 👻	Reset	Submit						
	Display:	25 - records	per page 🕕				_							
4564	records retu	urned (1 - 25 displ	aved) Click the	hyperl	inks for detailed inform	nation								
						idion.								
Page		3 4 5 6	789		Last									
			lso	late fie	ids A						MLST	Finety	ping ant	igens
id	isolate	aliases	country	year	disease	species	serogroup	Seqbin size (bp)	Contigs	ST	clonal complex	PorA VR1	PorA VR2	FetA VR
1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	2069108	364	4	ST-4 complex/subgroup IV	5-2	10	F1-5
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	4241338	652	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
120	F4698	NIBSC_2731; Z3515	Saudi Arabia	1987	carrier	Neisseria meningitidis	A	4192894	638	5	ST-5 complex/subgroup III	20	9	F3-1
369	M597	NIBSC_2781; Z4323	Israel	1988	invasive (unspecified/other)	Neisseria meningitidis	С	2068928	344	11	ST-11 complex/ET-37 complex	5	2-1	F5-5

Additional search terms can be combined using the '+' button. Sequence tag count queries will be combined with terms entered in other sections.

**Note:** Searches for 'all loci' with counts that include zero, e.g. 'count of any locus = 0' or with a '<' operator are not supported. This is because such searches have to identify every isolate for which one or more loci are not tagged. In databases with thousands of loci this can be a very expensive database query.

### 11.7.5 Query by tagged sequence status

Sequence tags identify the region of a contig within an isolate's sequence bin entries that correspond to a particular locus. The presence or absence of these tags can be queried as can whether or not the sequence has an a flag associated with. These flags designate specific characteristics of the sequences. Queries will be combined from the values entered in all form sections.

Make sure that the tagged sequences status fieldset is displayed by selecting it in the 'Modify form options' tab.

PubMLS	Query: Search   Profile/ST Breakdown: Isolate fields   Links: Contents   Home   O						
				Help 🗹	Toggle: 🚺 🛛 F	Field help: id	▼ Go
Search or	<sup>,</sup> browse Neisseri	a PubMLST da	atabas	9			
Enter search criteria or leave blank to browse all records. Modify form parameters t Isolate provenance/phenotype fields				Display/sort options		★ Modify form parameters	Modify form options
id	¥ = ¥	Enter value	+ 0	Order by: Id Display: 25 • records per page (	•	Click to add or remove additional query terms:	values

Select a specific locus in the dropdown box (or alternatively 'any locus') and a status. Available status values are:

- untagged
  - The locus has not been tagged within the sequence bin.
- tagged
  - The locus has been tagged within the sequence bin.
- complete
  - The locus sequence is complete.
- incomplete
  - The locus sequence is incomplete normally because it continues beyond the end of a contig.
- flagged: any
  - The sequence for the locus has a flag set.
- flagged: none
  - The sequence for the locus does not have a flag set.
- flagged: <specific flag>
  - The sequence for the locus has the specific flag chosen.

Control         Guery: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Help 🗹 Toggle: 🕄 Field help: id	Go
Search or browse Neisseria PubMLST database	
Enter search criteria or leave blank to browse all records.	Modify
Isolate provenance/phenotype fields Tagged sequence status	form
id v = v Enter value + 1 NEIS0001 (lpxC) v is flagged: internal stop codon v + 1	options
Display/sort options	
Order by: id • ascending • Reset Submit	
Display: 25	
1 record returned. Click the hyperlink for detailed information.	
Isolate fields 🔿 MLST Finetyping antigens	
id isolate aliases   country  year  disease   species   serogroup ST  clonal complex  PorA VR1  PorA VR2   FetA VR	
2077 153 alpha153 Germany 1999 carrier Neisseria meningitidis E 60 ST-60 complex 5 2 F1-7	
Analysis tools:	
Breakdown: Fields Two Field Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status	
Analysis: Codons Presence/Absence Genome Comparator BLAST	
Export: Dataset Contigs Sequences	

#### See also:

Sequence tag flags

## 11.7.6 Query by list of attributes

The query form can be modified with a list box in to which a list of values for a chosen attribute can be entered - this could be a list of ids, isolate names, alleles or scheme fields. This list will be combined with any other criteria or filter used on the page.

If the list box is not shown, add it by selecting it in the 'Modify form options' tab.

Query: Search   Profile/ST Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitio	ns	
	Help 🗗 🛛 Toggle: 🚺	Field help: id 🗸 🗸 Go
Search or browse Neisseria PubMLST databa	se	
Enter search criteria or leave blank to browse all records. Modify form parameters to	filter or enter a list of values.	Modify
	Display/sort options	* form
id	Order by: id	Modify form parameters options
	Display: 25 👻 records per page 🗊	Click to add or remove additional query terms:
		Provenance fields
		Allele designations/scheme field values
		Allele designation counts
		Allele designation status
		Tagged sequence counts
		Tagged sequence status
		Attribute values list
		Filters

Select the attribute to query and enter a list of values.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences do		Database submissions							
			Help 🗗	Toggi	e: 🚺 Field help: İd				Go
Search or browse Neisseria PubMLST data	abase								
Enter search criteria or leave blank to browse all records.									Modify
Isolate provenance/phenotype fields	A	ttribute values list							form
id 🗸 = 🗸 Entervalue +	🚺 Field	d: id							options
	1								
	2								
	4								
	5								
- Display/sort options		- Action	_						
Order by: id vascer	nding 👻	Reset Submi	t						
Display: 25 🗸 records per page 1									
5 records returned. Click the hyperlinks for detailed information.									
Isolate fields 1					MLST	Finet	yping antige	ens	
id isolate aliases country year disease		species		07			PorA VR2		
		species	serogroup	ST		PULA VILI			
1 A4/M1027 B1; NIBSC_2803; Z1001 USA 1937 invasive (unspeci	ified/other)	Neisseria meningitidis	A	4	ST-4 complex/subgroup IV	5-2	10	F1-5	
2 120M B35; NIBSC_2822; Z1035 Pakistan 1967 meningitis and se	ified/other) epticaemia	Neisseria meningitidis Neisseria meningitidis	A A	4 1		5-2 5-2	10	F1-5 F5-1	
2 120M B35; NIBSC_2822; Z1035 Pakistan 1967 meningitis and se 3 M00242905 UK 2000 invasive (unspeci	ified/other) epticaemia ified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B	4 1 1099	ST-4 complex/subgroup IV ST-1 complex/subgroup I/II	5-2			
2         120M         B35; NIBSC_2822; Z1035         Pakistan         1967         meningitis and se           3         M00242905         UK         2000         invasive (unspeci           4         M1027         B43; NIBSC_3076; Z1043         USA         1937         invasive (unspeci	ified/other) epticaemia ified/other) ified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A	4 1 1099 4	ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup IV	5-2 5-2 19	10 15		
2         120M         B35; NIBSC_2822; Z1035         Pakistan         1967         meningitis and se           3         M00242905         UK         2000         invasive (unspeci           4         M1027         B43; NIBSC_3076; Z1043         USA         1937         invasive (unspeci           5         M00240227         UK         2000         invasive (unspeci	ified/other) epticaemia ified/other) ified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A	4 1 1099 4	ST-4 complex/subgroup IV ST-1 complex/subgroup I/II	5-2 5-2	10		
2         120M         B35; NIBSC_2822; Z1035         Pakistan         1967         meningitis and se           3         M00242905         UK         2000         invasive (unspeci           4         M1027         B43; NIBSC_3076; Z1043         USA         1937         invasive (unspeci	ified/other) epticaemia ified/other) ified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A	4 1 1099 4	ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-4 complex/subgroup IV	5-2 5-2 19	10 15		
2         120M         B35; NIBSC_2822; Z1035         Pakistan         1967         meningitis and se           3         M00242905         UK         2000         invasive (unspeci           4         M1027         B43; NIBSC_3076; Z1043         USA         1937         invasive (unspeci           5         M00240227         UK         2000         invasive (unspeci	fied/other) epticaemia fied/other) fied/other) fied/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B	4 1 1099 4	ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I ST-32 complex/ET-5 complex	5-2 5-2 19	10 15		
2         120M         B35; NIBSC_2822; Z1035         Pakistan         1967         meningitis and see           3         M00242905         UK         2000         invasive (unspeci           4         M1027         B43; NIBSC_3076; Z1043         USA         1937         invasive (unspeci           5         M00240227         UK         2000         invasive (unspeci           Analysis tools:         UK         2000         invasive (unspeci	fied/other) epticaemia fied/other) fied/other) ified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B	4 1 1099 4 1100	ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I ST-32 complex/ET-5 complex	5-2 5-2 19	10 15		

## 11.7.7 Query filters

There are various filters that can additionally be applied to queries, or the filters can be applied solely on their own so that they filter the entire database.

Make sure that the filters fieldset is displayed by selecting it in the 'Modify form options' tab.

PubMLST		rofile/ST te fields   Scheme/alleles   Pu Home   Options   Profiles/seq		15				
					Help 🗗	Toggle: 🜖 🛛	Field help: id	▼ Go
Search or	browse Nei	isseria PubMLS	T databas	se				
Enter search criter	ria or leave blank to l	prowse all records. Modify form	n parameters to f	ilter or enter a list o	fvalues.			Modify
Isolate prover	nance/phenotype fiel	ds		— Display/sort opti	ons		×	form
id	▼ =	✓ Enter value	+ 0	Order by: id			Modify form parameters	options
				Display: 25	<ul> <li>records per page</li> </ul>	Ð	Click to add or remove additional query terms:	
							Provenance fields	
							Allele designations/scheme field v	values
							Allele designation counts	
							O Allele designation status	
							Contract Tagged sequence counts	
							Tagged sequence status	
							Attribute values list	
							Filters	
l								

The filters displayed will depend on the database and what has been defined within it. Common filters are:

- Publication Select one or more publication that has been linked to isolate records.
- Project Select one or more project that isolates belong to.

- Profile completion This is commonly displayed for MLST schemes. Available options are:
  - complete All loci of the scheme have alleles designated.
  - incomplete One or more loci have not yet been designated.
  - partial The scheme is incomplete, but at least one locus has an allele designated.
  - started At least one locus has an allele designated. The scheme mat be complete or partial.
  - not started The scheme has no loci with alleles designated.
- Sequence bin Specify whether any sequence data has been associated with a record. Specific threshold values may be selected if these have been *set up for the database*.
- Provenance fields Dropdown list boxes of values for specific provenance fields may be present if set for the database. Users can choose to *add additional filters*.

# 11.8 Querying by allelic profile

If a scheme, such as MLST, has been defined for an isolate database it is possible to query the database against complete or partial allelic profiles. Even if no scheme is defined, queries can be made against all loci. This can also be done in sequence definition databases if the scheme has a primary key field defined.

On the index page, click 'Search by combinations of loci (profiles)' for any defined scheme. Enter either a partial (any combination of loci) or complete profile.

Publications Problem Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently if sample. Query database Search or browse database Search by combinations of loci (profiles) Projects - main projects defined in database. Option settings Set display and query options for locus, schemes or scheme fields.	t should be noted that it does not represent a population  Submissions  Manage submissions
Breakdown       Export       Analysis         Single field       • Export dataset       • Contigs       • Contigs         • Unique combinations       • Scheme and alleles       • Sequences - XMFA / concatenated FASTA formats       • Codon usage         • Publications       • Sequence bin       • Sequences - XMFA / concatenated FASTA formats       • BLAST	Miscellaneous     • Description of database fields

If multiple schemes are defined, you may have to select the scheme you wish to query in the 'Schemes' dropdown box and click 'Select'.

Query: Search   Browse   Profile/ST Breakdown: Isolate fields   Scheme Links: Contents   Home   Options		
		Toggle: [
Search Neisseria PubMLST da	tabase by combinations of loci	
Schemes		
Please select the scheme you would like to query:		
MLST - S	elect	
Please enter your allelic profile below. Blank loci wil     abcZ adk aroE fumC gdh		
	pdhC pgm ST:	Autofill
Filters	Options	- Display/sort options
Project:	i Search: Exact or nearest match	Order by: id     ascending
Include old record versions		Display: 25 🗸 records per page 👔
Reset Submit		

Enter the combination of alleles that you want to query for. Fields can be left blank.

Construction Const	
	Toggle: i
Search Neisseria PubMLST database by combinations of loci	
Schemes	
Please select the scheme you would like to query:	
MLST - Select	
Please enter your allelic profile below. Blank loci will be ignored. — Autofill profile by searching remote database —	
abcZ adk aroE fumC gdh pdhC pgm ST: Autofill	
Filters     Options     Display/sort options       Project: <ul> <li>I Search: Exact or nearest match</li> <li>Order by: id</li> <li>ascending</li> </ul>	•
□ Include old record versions Display. 25	
Action	
Reset Submit	

Alternatively, for scheme profiles, you can enter a primary key value (e.g. ST) and select 'Autofill' to automatically fill in the associated profile.

Query:         Search   Browse   Profile/ST   List           Breakdown:         Isolate fields   Scheme/alleles   Publications           Links:         Contents   Home   Options   Profiles/sequences definitions   Database submissions	
	Toggle: i
Search Neisseria PubMLST database by combinations of loci	
Schemes	
Please select the scheme you would like to query:	
MLST V Select	
abcZ     adk     aroE     fumC     gdh     pdhC     pgm     ST: 44     Autofill       9     6     9     9     6     9	
- Filters - Options - Display/sort options	
Project:	-
□ Include old record versions Display: 25	
Action	
Reset Submit	

Select the number of loci that you'd like to match in the options dropdown box. Available options are:

- Exact or nearest match
- Exact match only
- x or more matches
- y or more matches
- z or more matches

Where x,y, and z will range from n-1 to 1 where n is the number of loci in the scheme.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
	Toggle: i
Search Neisseria PubMLST database by combinations of loci	
Schemes	
Please select the scheme you would like to query:	
MLST v Select	
Please enter your allelic profile below. Blank loci will be ignored. — Autofill profile by searching remote datab	ISE
abcZ adk aroE fumC gdh pdhC pgm ST: 44	utofill
Filters Options Display/sort of	tions
Project: View Search: Exact or nearest match View Order by: id	
Include old record versions     Display: 2	records per page i
Action	
Reset Submit	

Click 'Submit'.

											-
											Toggle
ar	ch Neis	sseria P	ubMLST d	atal	base by combin	nations of loc	i				
her	nes										
ase	select the se	cheme you wou	Id like to query:								
LST				Selec	t						
					_						
Ple	ase enter vo	our allelic profile	below. Blank loci v	will be	ignored.	-Autofill profile by search	nina remote	database			
	cZ ad			idh	pdhC pgm	ST: 44		Autofill			
	9 6			9	6 9			Addini			
- Filt	ore				Options			/sort options			
	Project					or nearest match 👻		by: id		- ascend	na –
					I Search. Exact	or nearest match +	Order	uy. Iu		<ul> <li>ascenu</li> </ul>	ng 👻
Act	ion	ecord versions					Disp	lay: 25 → records per page 👔	]		
Act Re:	ion	mit					Disp	Ilay: 25 → records per page [	]		
Act Re:	ion set Sub atches found	mit 1 (7 loci).	ed). Click the hype	rlinks f	or detailed information.		Disp	Ilay: 25 → records per page [	]		
- Act Re: ct m: recc	ion set Sub atches found ords returned	mit 1 (7 loci). 1 (1 - 25 display		rlinks f	or detailed information.		Disp	Ilay: 25 → records per page [	]		
- Act Re: ct ma reco	ion set Sub atches found ords returned	mit 1 (7 loci).	ed). Click the hype	rlinks f	or detailed information.		Disp	ılay: 25 → records per page [	]		
- Act Re: ct ma reco	ion set Sub atches found ords returned	mit 1 (7 loci). 1 (1 - 25 display					Disp			voing apfi	ANS
- Act Re: ct m: recc je:	ion set Sub atches found ords returned 1 2 3	mit 1 (7 loci). 1 (1 - 25 display	Last	Isolat	e fields 🖪	species		MLST	Finet	yping antig	
- Act Re: ct m: recc je: (	ion set Sub atches found ords returned	mit d (7 loci). d (1 - 25 display ) ( 4 ( 5 ) ≥				species Neisseria meningitidis	Disp serogroup B	MLST	Finet PorA VR1	yping antig PorA VR2	
- Act Re: ct m: recc ge: (	ion set Sub atches found ords returned 1 2 3 isolate	mit d (7 loci). d (1 - 25 display ) ( 4 ( 5 ) ≥	Last country	Isolat year	e fields 👔 disease		serogroup	ML.ST ST clonal complex	Finet PorA VR1		
Act Res ct ma recc ge: ( 11 '0 27	ion set Sub atches found ords returned 1 2 3 isolate 19 38 NG E30	mit d (7 loci). d (1 - 25 display ) ( 4 ( 5 ) ≥	Last Country Germany	Isolat year 1999	e fields 👔 disease carrier	Neisseria meningitidis	serogroup B NG B	MLST ST clonal complex 44 ST-41/44 complex/Lineage 3 44 ST-41/44 complex/Lineage 3	Finet PorA VR1 21		
- Act Re: ct m: recc ge: ( 11 0 27 74	ion set Sub atches found ords returned 1 2 3 isolate 19 38 NG E30 99 182	mit d (7 loci). d (1 - 25 display ) 4 5 > aliases	Country Germany Germany Norway Canada	Isolat year 1999 1999 1988 1999	e fields 👔 disease carrier carrier carrier invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	serogroup B NG B B	MLST ST clonal complex 44 ST-41/44 complex/Lineage 3 44 ST-41/44 complex/Lineage 3 44 ST-41/44 complex/Lineage 3 44 ST-41/44 complex/Lineage 3	Finet PorA VR1 21	PorA VR2	FetA VR
- Act Re: ct m: recc ge: ( 11 70 277 74 92	ion set Sub atches found ords returned 1 2 3 isolate 19 38 NG E30 99 182 99-132	mit d (7 loci). d (1 - 25 display ) 4 5 > aliases	Country Germany Germany Norway Canada Canada	Isolat year 1999 1999 1988 1999 1999	e fields 👔 disease carrier carrier carrier invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	serogroup B NG B B B B	MLST ST clonal complex 44 ST-41/44 complex/Lineage 3 44 ST-41/44 complex/Lineage 3 44 ST-41/44 complex/Lineage 3 44 ST-41/44 complex/Lineage 3	Finet PorA VR1 21	PorA VR2	FetA VR F1-7
Act Re: recc ge: ( 41 70 227 74 92 75	ion set Sub atches found ords returned 1 2 3 isolate 19 38 NG E30 99 H32 99-H32 99-H32 0069/93	mit d (7 loci). d (1 - 25 display ) 4 5 > aliases Z4692	Country Germany Germany Norway Canada Canada Czech Republic	Isolat year 1999 1999 1988 1999 1999 1993	e fields 👔 disease carrier carrier carrier invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	serogroup B NG B B B NG NG	MLST           ST         clonal complex/Lineage 3           44         ST-41/44 complex/Lineage 3	Finet PorA VR1 21 22	PorA VR2 16 14-4	FetA VR
Act Re: act m: ) recc ge: ( 41 41 70 427 774 792 975 097	ion set Sub atches found ords returned 1 2 3 isolate 19 38 NG E30 99-132 99-132 0069/93 3532	mit d (7 loci). d (1 - 25 display ) 4 5 > aliases	Country Germany Germany Norway Canada Canada Czech Republic The Netherlands	Isolat year 1999 1999 1988 1999 1999 1993 1975	e fields 2 disease carrier carrier invasive (unspecified/other) invasive (unspecified/other) carrier	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	serogroup B NG B B B B S G B S	MLST 25T clonal complex 44 ST-41/44 complex/Lineage 3 44 ST-41/44 complex/Lineage 3	Finet PorA VR1 21 22	PorA VR2 16 14-4 4	FetA VR F1-7 F1-7
Act Re: act m: ) recc ge: ( 41 70 127 774 792 075	ion set Sub atches found ords returned 1 2 3 isolate 19 38 NG E30 99 H32 99-H32 99-H32 0069/93	mit d (7 loci). d (1 - 25 display ) 4 5 > aliases Z4692	Country Germany Germany Norway Canada Canada Czech Republic	Isolat 999 1999 1988 1999 1993 1993 1975 1993	e fields 👔 disease carrier carrier carrier invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	serogroup B NG B B B NG NG	MLST           ST         clonal complex/Lineage 3           44         ST-41/44 complex/Lineage 3	Finet PorA VR1 21 22 22	PorA VR2 16 14-4	FetA VR F1-7

# 11.9 Retrieving isolates by linked publication

Click 'Publications' in the Breakdown section of the contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Neisseria PubMLST database	
The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently is sample.	
Query database <ul> <li>Search or browse database</li> <li>Search by combinations of loci (profiles)</li> <li>Projects - main projects defined in database.</li> </ul> <ul> <li>Option settings</li> <li>Set general options - including isolate table field handling.</li> <li>Set display and query options for locus, schemes or scheme fields.</li> </ul>	Submissions Manage submissions General information Isolates: 35423 Last updated: 2015-08-27 Update history About BIGSdb
Breakdown       Export       Analysis         • Single field       • Export dataset       • Contigs         • Two field       • Contigs       • Contigs         • Scheme and alleles       • Sequences - XMFA / concatenated FASTA formats       • Codon usage         • Publications       • Sequences in       • BLAST	Miscellaneous     • Description of database fields loci

A list of publications linked by isolates within the database will be displayed.

PubML	ST	Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submis	ions	
Publicat	tior	h breakdown of dataset		
	ery by	Display	Action	
Auth	or: A	Il authors	Submit	
Ye	ar: A	ll years		
Page: 1 2	: 3			Isolates in
id	Year	Citation	Title	database
17517841 2				
		Taha MK, Vázquez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, Cafferkey MT, Carion F, Christensen JJ, Diggle M, Edwards G, Enríquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen I Levenet I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother <b>51</b> : 2784-92	Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis.	1670 isolates
	2008	F, Christensen JJ, Diggle M, Edwards G, Enríquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen I Levenet I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki	susceptibility of Neisseria meningitidis.	1670 isolates
18815379 2	2008	F, Christensen JJ, Diggle M, Edwards G, Enríquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen I Levenet I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother 51: 2784-92 Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Nal	susceptibility of Neisseria meningitidis. Role of selection in the emergence of lineages and the evolution of	
18815379 2 15776372 2	2008 2005 2004	F, Christensen JJ, Diggle M, Edwards G, Enriquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen I Levrent I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanellin F, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother 51: 2784-92 Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Naf Acad Sci U S A 105: 15082-7 Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M,	susceptibility of Neisseria meningitidis. Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis. Genetic analysis of meningococci carried by children and young	1054 isolates
18815379 2 15776372 2 15528708 2	2008 2005 2004 2007	F, Christensen JJ, Diggle M, Edwards G, Enriquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen I Levrent I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother 51: 2784-92 Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Naf Acad Sci U S 4105: 15082-7 Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M, Vogel U (2005) J Infect Dis 191: 1263-71 Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Caugant DA, Maiden MC (2004) J Clin Microbiol 42:	susceptibility of Neisseria meningitidis. Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis. Genetic analysis of meningococci carried by children and young adults. Distribution of serogroups and genotypes among disease- associated and carried isolates of Neisseria meningitidis from the	1054 isolates 822 isolates
18815379 2 15776372 2 15528708 2 17825091 2	2008 2005 2004 2007	F, Christensen JJ, Diggle M, Edwards G, Enríquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen I Levrent I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother 51: 2784-92 Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Nal Acad Sci U S A 105: 15082-7 Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M, Vogel U (2005) J Infect Dis 191: 1263-71 Yazdankhab SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Caugant DA, Maiden MC (2004) J Clin Microbiol 42: 5146-53 Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007)	susceptibility of Neisseria meningitidis. Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis. Genetic analysis of meningococci carried by children and young adults. Distribution of serogroups and genotypes among disease- associated and carried isolates of Neisseria meningitidis from the Czech Republic, Greece, and Norway. Species status of Neisseria gonorrhoeae: evolutionary and	1054 isolates       822 isolates       667 isolates

These can be filtered by author and/or year, and the sort order changed.

ublica	ntio	n breakdown of dataset		
	hor:	Olsplay     Action       Iolley KA     ▼       Order by:     number of isolates ▼       descending ▼     Submit       Submit     Display:       25 ▼     records per page		
records re PubMed id	eturne Year	1. Click the hyperlinks for detailed information. Citation	Title	Isolates in database
	2007	Taha MK, Vázquez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, Cafferkey MT, Carion F, Christensen JJ, Diggle M, Edwards G, Enriquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen L, Levenet I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaß G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother <b>51</b> : 2784-92	Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis.	1670 isolates
8815379	2008	Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Natl Acad Sci U S A 105: 15082-7	Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis.	1054 isolates
5776372	2005	Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M, Vogel U (2005) J Infect Dis <b>191:</b> 1263-71	Genetic analysis of meningococci carried by children and young adults.	822 isolates
5528708	2004	Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Caugant DA, Maiden MC (2004) J Clin Microbiol <b>42:</b> 5146-53	Distribution of serogroups and genotypes among disease-associated and carried isolates of Neisseria meningitidis from the Czech Republic, Greece, and Norway.	667 isolates
7825091	2007	Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007) BMC Biol 5: 35	Species status of Neisseria gonorrhoeae: evolutionary and epidemiological inferences from	576 isolates
			multilocus sequence typing.	

To display the isolate records for any of the displayed publications, click the button to the right of the citation.

<u>िर्यतेल</u> ्य	LST	Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions		
Publica	atio	breakdown of dataset		
	thor:	Olley KA     ✓     Order by: number of isolates ▼ descending ▼     Submit       Ill years ▼     Display: 25 ▼ records per page		
13 records r	eturne	I. Click the hyperlinks for detailed information.		
PubMed id	Year	Citation	Title	lsolates in database
17517841	2007	Taha MK, Vázquez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, Cafferkey MT, Carion F, Christensen JJ, Diggle M, Edwards G, Enríquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen L, Levenet I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother 51: 2784-92	Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis.	1670 isolates
18815379	2008	Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Natl Acad Sci U S A 105: 15082-7	Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis.	1054 isolates
15776372	2005	Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M, Vogel U (2005) J Infect Dis <b>191:</b> 1263-71	Genetic analysis of meningococci carried by children and young adults.	822 isolates
15528708	2004	Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Caugant DA, Maiden MC (2004) J Clin Microbiol <b>42:</b> 5146-53	Distribution of serogroups and genotypes among disease-associated and carried isolates of Neisseria meningitidis from the Czech Republic, Greece, and Norway.	667 isolates
17825091	2007	Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007) BMC Biol 5: 35	Species status of Neisseria gonorrhoeae: evolutionary and epidemiological inferences from multilocus sequence typing.	576 isolates
15537808	2005	Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005) Mol Biol Evol <b>22:</b> 562-9	The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in Neisseria meningitidis.	378 isolates
		Bennett JS, Griffiths DT, McCarthy ND, Sleeman KL, Jolley KA, Crook DW, Maiden MC (2005) Infect Immun 73: 2424-32	Genetic diversity and carriage dynamics of Neisseria lactamica in infants.	271 isolates
11101585	2000	Iollev KA Kalmusova I Feil F I Gunta S Musilek M Kriz P Maiden MC (2000) I Clin Microbiol 38: 4492-8	Carried meningococci in the Czech Republic: a	017:

The abstract of the paper will be displayed (if available), along with all isolates linked to it.

ti	on que		d in the ed id: 178		sseria PubMLS	T database						
n		ry (PubM	ed id: 178	2500								
n		ry (PubM	ed id: 178	2500								
	ott IC Tolly			2005	91)							
					art CA, Feavers IM, Maiden M onary and epidemiological in			o tuni	na			
									typing method. In addition, th	ese data pe	ermitted ph	vlogenetic and
									on of these data demonstrate			
									d determining whether mixed			
			-						-			
į	cords retui	med (1 - 25 dis	splayed). Click	the hy	perlinks for detailed informat	ion.						
	1 2	3 4 5	6 7 8	9	> Last							
				-1								
T	isolate				solate fields 👔				MLST	Finety	ping antig	ens
		aliases	country	year	solate fields <mark>i disease</mark>	species	serogroup	ST		Finety PorA VR1	/ping antig PorA VR2	
	A4/M1027	aliases B1; Z1001	country USA	year		•	serogroup A	ST 4				
				year 1937	disease	Neisseria meningitidis			clonal complex	PorA VR1	PorA VR2	FetA VR
	A4/M1027	B1; Z1001	USA	year 1937 1967	disease invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A	4	clonal complex ST-4 complex/subgroup IV	PorA VR1 5-2	PorA VR2 10	FetA VR F1-5
	A4/M1027 120M	B1; Z1001 B35; Z1035	USA Pakistan	year 1937 1967 1975	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A	4 1	cional complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-5 complex/subgroup III	PorA VR1 5-2 5-2	PorA VR2 10 10	FetA VR F1-5 F5-1
	A4/M1027 120M 7891	B1; Z1001 B35; Z1035 B54; Z1054	USA Pakistan Finland Canada	year 1937 1967 1975 1971	disease invasive (unspecified/other) meningitis and septicaemia	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A	4 1 5	cional complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II	PorA VR1 5-2 5-2 20	PorA VR2 10 10 9	FetA VR F1-5 F5-1 F3-1
	A4/M1027 120M 7891 6748 129	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092	USA Pakistan Finland Canada Germany	year 1937 1967 1975 1971 1964	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A	4 1 5 1	cional complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I	PorA VR1 5-2 5-2 20 18-1 5-2	PorA VR2 10 10 9 3	FetA VR F1-5 F5-1 F3-1 F5-1
	A4/M1027 120M 7891 6748	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073	USA Pakistan Finland Canada Germany	year 1937 1967 1975 1971 1964 1968	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A	4 1 5 1 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/II ST-5 complex/subgroup III ST-1 complex/subgroup I/II	PorA VR1 5-2 5-2 20 18-1	PorA VR2 10 10 9 3 10	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6
	A4/M1027 120M 7891 6748 129 139M	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099	USA Pakistan Finland Canada Germany Philippines Ghana	year 1937 1967 1975 1971 1964 1968 1973	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A A	4 1 5 1 1 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup IV	PorA VR1 5-2 20 18-1 5-2 5-2 5-2	PorA VR2 10 9 3 10 10	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6 F5-1
	A4/M1027 120M 7891 6748 129 139M \$3131	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099 B213; Z1213 B227; Z1227	USA Pakistan Finland Canada Germany Philippines Ghana Denmark	year 1937 1967 1975 1971 1964 1968 1973 1974	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A A A A A	4 1 5 1 1 1 4	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup IV ST-5 complex/subgroup II	PorA VR1 5-2 20 18-1 5-2 5-2 7 5-2 7 5-1	PorA VR2 10 9 3 10 10 10 13-1	FetA VR F1-5 F5-1 F3-1 F5-1 F3-6 F5-1 F1-5
	A4/M1027 120M 7891 6748 129 139M S3131 S4355	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099 B213; Z1213 B227; Z1227 B269; Z1269	USA Pakistan Finland Canada Germany Philippines Ghana Denmark Burkina Faso	year 1937 1967 1975 1971 1964 1968 1973 1974 1963	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A A A	4 5 1 1 1 4 5	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup IV ST-5 complex/subgroup IV ST-4 complex/subgroup IV	PorA VR1 5-2 20 18-1 5-2 5-2 5-2 7	PorA VR2 10 9 3 10 10 10 13-1 9	FetA VR F1-5 F5-1 F3-1 F3-6 F5-1 F5-1 F1-5 F3-1
	A4/M1027 120M 7891 6748 129 139M S3131 S4355 10	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099 B213; Z1213 B227; Z1227	USA Pakistan Finland Canada Germany Philippines Ghana Denmark	year 1937 1967 1975 1971 1964 1968 1973 1974 1963 1963	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A A A A A A A A	4 1 5 1 1 4 5 4	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup IV ST-5 complex/subgroup IV ST-4 complex/subgroup IV ST-4 complex/subgroup I/I	PorA VR1 5-2 20 18-1 5-2 5-2 7 5-2 7 5-1 7	PorA VR2 10 9 3 10 10 13-1 9 13-1	FetA VR F1-5 F5-1 F3-1 F3-6 F5-1 F3-6 F5-1 F1-5 F3-1 F1-5
	A4/M1027 120M 7891 6748 129 139M S3131 S4355 10 20 26	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099 B213; Z1213 B227; Z1227 B269; Z1269 B275; Z1278	USA Pakistan Finland Canada Germany Philippines Ghana Denmark Burkina Faso Niger	year 1937 1967 1975 1971 1964 1968 1973 1974 1963 1963 1963	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A A A A A A A A A	4 1 5 1 1 4 5 4 1 4	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup IV ST-5 complex/subgroup IV ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I	PorA VR1 5-2 20 18-1 5-2 5-2 7 5-1 7 5-1 7 5-2 7	PorA VR2 10 9 3 10 10 13-1 9 13-1 10 13	FetA VR F1-5 F5-1 F3-1 F3-1 F3-6 F5-1 F3-6 F5-1 F1-5 F1-5 F1-7 F1-7 F1-5
	A4/M1027 120M 7891 6748 129 139M S3131 S4355 10 20 26 255	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099 B213; Z1213 B227; Z1227 B269; Z1269 B275; Z1275 B278; Z1278 B318; Z1318	USA Pakistan Finland Canada Germany Philippines Ghana Denmark Burkina Faso Niger Burkina Faso	year 1937 1967 1975 1971 1964 1968 1973 1974 1963 1963 1963 1966	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A A A A A A A A A A A A A	4 5 1 1 4 5 4 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I	PorA VR1 5-2 20 18-1 5-2 5-2 7 5-2 7 5-1 7 5-1 7 5-2 7 7 7-2	PorA VR2 10 10 9 3 10 10 13-1 9 13-1 10 13-1 10 13 13-1	FetA VR F1-5 F5-1 F3-1 F3-6 F5-1 F1-5 F3-1 F1-5 F1-7 F1-5 F1-5 F1-5 F1-5 F1-5
	A4/M1027 120M 7891 6748 129 139M S3131 S4355 10 20 26 255 243	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099 B213; Z1213 B227; Z1227 B269; Z1269 B275; Z1275 B278; Z1278 B318; Z1318 B3362; Z1362	USA Pakistan Finland Germany Philippines Ghana Denmark Burkina Faso Niger Burkina Faso Cameroon	year 1937 1967 1975 1971 1964 1968 1973 1974 1963 1963 1963	disease invasive (unspecified/other) meningtis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A A A A A A A A A A A A A A A A	4 1 5 1 1 4 5 4 1 4 4 4	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup IV ST-5 complex/subgroup IV ST-4 complex/subgroup IV ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I	PorA VR1 5-2 5-2 20 18-1 5-2 5-2 7 5-1 7 5-1 7 5-2 7 7 7 7-2 7	PorA VR2 10 9 3 10 10 13-1 9 13-1 10 13	FetA VR F1-5 F5-1 F3-1 F3-1 F3-6 F5-1 F3-6 F5-1 F1-5 F1-5 F1-7 F1-7 F1-5
	A4/M1027 120M 7891 6748 129 139M \$3131 \$4355 10 20 26 255 243 393	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099 B213; Z1213 B227; Z1227 B269; Z1269 B275; Z1275 B278; Z1278 B318; Z1318 B362; Z1362 B392; Z1362	USA Pakistan Finland Canada Germany Philippines Ghana Denmark Burkina Faso Niger Niger Burkina Faso Cameroon Greece	year 1937 1967 1975 1971 1964 1968 1973 1963 1963 1963 1966 1966 1968	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A A A A A A A A A A A A A A A A	4 1 5 1 1 4 5 4 1 4 4 4 4 4 1	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup I/I	PorA VR1 5-2 5-2 20 18-1 5-2 5-2 7 5-1 7 5-2 7 7 7-2 7 7 2 5-2 7	PorA VR2 10 10 9 3 10 10 13-1 9 13-1 10 13 13 13-1 13 10	FetA VR F1-5 F5-1 F3-1 F5-1 F5-1 F1-5 F3-1 F1-5 F1-7 F1-5 F1-5 F1-5 F1-5 F1-5 F5-1
	A4/M1027 120M 7891 6748 129 139M S3131 S4355 10 20 26 255 243	B1; Z1001 B35; Z1035 B54; Z1054 B73; Z1073 B92; Z1092 B99; Z1099 B213; Z1213 B227; Z1227 B269; Z1269 B275; Z1275 B278; Z1278 B318; Z1318 B3362; Z1362	USA Pakistan Finland Canada Germany Philippines Ghana Denmark Burkina Faso Niger Niger Burkina Faso Cameroon Greece Djibouti	year 1937 1967 1975 1971 1964 1968 1973 1974 1963 1963 1963 1966 1966 1968 1966	disease invasive (unspecified/other) meningtis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A A A A A A A A A A A A A A A A	4 1 5 1 1 4 5 4 1 4 4 4 4 4	clonal complex ST-4 complex/subgroup IV ST-1 complex/subgroup I/I ST-5 complex/subgroup I/I ST-1 complex/subgroup I/I ST-1 complex/subgroup I/I ST-4 complex/subgroup IV ST-5 complex/subgroup IV ST-4 complex/subgroup IV ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I ST-4 complex/subgroup I/I	PorA VR1 5-2 5-2 20 18-1 5-2 5-2 7 5-1 7 5-1 7 5-2 7 7 7 7-2 7	PorA VR2 10 9 3 10 10 13-1 13-1 10 13-1 13-1 13-1 13-	FetA VR F1-5 F5-1 F3-1 F5-1 F5-1 F1-5 F1-5 F1-7 F1-5 F1-7 F1-5 F1-5 F1-5 F1-5 F1-5

# 11.10 User-configurable options

The BIGSdb user interface is configurable in a number of ways. Choices made are remembered between sessions. If the database requires you to log on, the options are associated with your user account, whereas if it is a public database, that you haven't logged in to, the options are associated with a browser cookie so they will be remembered if you connect from the same computer (using the same browser).

Most options are set by clicking the 'Set general options' link on the database contents page. Most of the available options are visible for isolate databases, whereas sequence definition databases have fewer available.

Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it sample.	
Query database     Search or browse database     Search by combinations of loci (profiles)     Projects - main projects defined in database.     Option settings     Set general options - including isolate table field handling.     Set display and query options for locus, schemes or scheme fields.	Submissions  Manage submissions  General information  Isolates: 35423  Last updated: 2015-08-27  Update history  About BIGSdb
Breakdown       Export       Analysis         • Single field       • Export dataset       • Contigs         • Unique combinations       • Scheme and alleles       • Sequences - XMFA / concatenated FASTA formats       • Codon usage         • Publications       • Sequence bin       • Sequences - XMFA / concatenated FASTA formats       • Export dataset	Miscellaneous     • Description of database fields

## 11.10.1 General options

The general options tab is displayed by default. If another tab is being shown, click the 'General options' header.

PubMLST/ Brea	ry: Search   Browse   Profile/ST   List akdown: Isolate fields   Scheme/alleles   Publications s: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Set database o	options
	for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set ry page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.
✓ General options	
Display 100 👻 Display locus al	top and bottom 👻 ucleotides per line in sequence alignments. nucleotides of flanking sequence (where available).
<ul> <li>Main results table</li> </ul>	
<ul> <li>Isolate record displa</li> </ul>	ay
<ul> <li>Provenance field dis</li> </ul>	play
<ul> <li>Query filters</li> </ul>	
Reset Click the reset button to re Reset all to defaults	emove all user settings for this database - this includes locus and scheme field preferences.

The general tab allows the following options to be modified:

• Records per page

- Page bar position
- Nucleotides per line Some analyses display sequence alignments. This option allows you to set the width of these alignments so suit your display.
- Flanking sequence length This sets the length of flanking sequence upstream and downstream of a particular locus that is included whenever a sequence is displayed. Flanking sequences are displayed fainter that the locus sequence.
- Locus aliases Loci can have multiple names (aliases). Setting this option will display all alternative names in results tables.
- Tooltips (beginner's mode) Most query forms have help available in the form of information tooltips. These can be switched on/off here. They can also be toggled off by clicking the Toggle: 'i' button at the top-right of the display of some pages.

Click 'Set options' to remember any changes you make.

### 11.10.2 Main results table

The 'main results table' tab contains options for the display of paged results following a query.

Click the 'Main results table' header to display the tab.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Set database options
Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.
▼ General options
Display 25
➤ Main results table
Isolate record display
Provenance field display
► Query filters
Reset Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences. Reset all to defaults

The 'main results table' tab will scroll up.

Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Help C
Set database options
Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.
General options
✓ Main results table
Hyperlink allele designations where possible.  Display information about sequence bin records tagged with locus information (tooltip).  Display sequence bin size.  Display contig count.  Display publications.  Set options
► Isolate record display
Provenance field display
Query filters
Reset Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences. Reset all to defaults

This tab allows the following options to be modified:

- Hyperlink allele designations Hyperlinks point to an information page about the particular allele definition. Depending on the locus, these may exist on a different website.
- Differentiate provisional allele designations Allele designations can be set as confirmed or provisional, usually depending on the method of assignment. Selecting this option will display provisional designations in a different colour to confirmed designations.
- Information about sequence bin records Creates a tooltip that displays details about sequence tags corresponding to a locus.
- Sequence bin records Displays a tooltip linking to the sequence tag if available.
- Sequence bin size Displays the size of the sum of all contigs associated with each isolate record.
- Contig count Displays the number of contigs associated with each isolate record.
- Publications Displays citations with links to PubMed for each record.

### 11.10.3 Isolate record display

The 'isolate record display' tab contains options for the display of a full isolate record.

Click the 'Isolate record display' tab to display the tab.

Query:         Search   Browse   Profile/ST   List           Breakdown:         Isolate fields   Scheme/alleles   Publications           Links:         Contents   Home   Options   Profiles/sequences definitions   Database submissions
Set database options
Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.
- General options
Display 25 • records per page. Page bar position: top and bottom • Display 100 • nucleotides per line in sequence alignments. Display 100 • nucleotides of flanking sequence (where available). Ø Display locus aliases if set. Ø Enable tooltips (beginner's mode). Set options
► Main results table
► Isolate record display
Provenance field display
Query filters
Reset Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences. Reset all to defaults

### The 'Isolate record display' tab will scroll up.

Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Set database options
Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.
General options
Main results table
✓ Isolate record display
Differentiate provisional allele designations.     Display sender, curator and last updated details for allele designations (tooltip).     Display information about sequence bin records tagged with locus information (tooltip).     Display information about whether alleles have flags defined in sequence definition database (shown in sequence detail tooltip).     Display full information about sample records (tooltip).     Set options
Provenance field display
Query filters
Reset Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences. Reset all to defaults

This tab allows the following options to be modified:

• Differentiate provisional allele designations - Allele designations can be set as confirmed or provisional, usually depending on the method of assignment. Selecting this option will display provisional designations in a different colour to confirmed designations.

- Display sender, curator and last updated records Displays a tooltip containing sender information next to each allele designation.
- Sequence bin information Displays a tooltip with information about the position of the sequence if tagged within the sequence bin.
- Allele flags Displays information about whether alleles have flags defined in sequence definition databases.
- Display full information about sample records Used when the database is used as part of a basic laboratory information management system (LIMS). This option will display records of samples available for the displayed isolate.

### 11.10.4 Provenance field display

The 'provenance field display' tab contains checkboxes for fields to display in the main results table.

Click the 'Provenance field display' tab to display the tab.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Set database options
Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.
▼ General options
Display 25 • records per page. Page bar position: top and bottom • Display 100 • nucleotides per line in sequence alignments. Display 100 • nucleotides of flanking sequence (where available). Ø Display locus aliases if set. Ø Enable tooltips (beginner's mode). Set options
Main results table
► Isolate record display
Provenance field display
Query filters
Reset
Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.
Reset all to defaults

The 'Provenance field display' tab will scroll up.

Links: Contents   Hor Set database options Here you can set options for your use of the	elds   Schemeralleles   Publications ne   Options   Profiles/sequences definitions   I website. Options are remembered between s hing the page (Shift + Refresh) as some pages	essions and affect the current database (Neisseria	a PubMLST) only. If some of the options don't appear to set
General options     Main results table     Isolate record display			
	and then selecting the 'Customize' option.	table following a query. Settings for displaying loc amoxicillin sulphonamide cettriaxone eftriaxone_range chloramphenicol chloramphenicol_range cefotaxime cefotaxime_range ffampicin iffampicin_range ciprofloxacin	us and scheme data can be made by performing a ciprofloxacin_range pending_assembly assembly_status ENA_accession private_project comments sender curator date_entered date_stamp
Query filters  Reset  Click the reset button to remove all user so Reset all to defaults	ettings for this database - this includes locus a	nd scheme field preferences.	

Some fields will be checked by default - these are defined during *database setup* (maindisplay option).

Check any fields that you wish to be displayed and then click 'Set options'. You can return to the default selection by clicking 'Default' followed by 'Set options'.

### 11.10.5 Query filters

The 'query filters' tab contains checkboxes for provenance fields and scheme completion status. Checking these results in drop-down list box filters appearing in the query page *filters fieldset*.

Click the 'Query filters' tab to display the tab.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Set database options
Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.
General options
Display 25 • records per page.
Page bar position: top and bottom
Display 100 🔹 nucleotides of flanking sequence (where available).
✓ Display locus aliases if set. ✓ Enable tooltips (beginner's mode).
Set options
Main results table
► Isolate record display
Provenance field display
► Query filters
Reset
Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.
Reset all to defaults

### The 'Query filters' tab will scroll up.

solate record display		
Provenance field display		
Query filters		
Select the fields for which you would like dropdow isolate country continent region year epidemiological_year age_yrt age_mth sex disease source epidemiology species serogroup MLEE_designation serotype ET_no penicillin penicillin penicillin sulphonamide ceftriaxone ceftriaxone_range chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chloramphenicol chlo	n lists containing known values on which to filter query results. These w ENA_accession private_project comments sender curator date_entered datestamp MLST profile completion finetyping antigens profile completion ADP-heptose biosynthesis profile completion Antipacy-HRNA biosynthesis profile completion Antipacy enses profile completion Capsule Region A - Serogroup A profile completion Capsule Region A - Serogroup B profile completion Capsule Region A - Serogroup B profile completion Capsule Region A - Serogroup D profile completion Capsule Region A - Serogroup D profile completion Capsule Region A - Serogroup D profile completion Capsule Region A - Serogroup L profile completion Capsule Region A - Serogroup L profile completion Capsule Region A - Serogroup Y profile completion Capsule Region C - Serogroup Y profile completion Capsule Region D profile completion Capsule Region D profile completion Capsule Region D profile completion Capsule Region D profile completion Conjugative Plasmid profile completion Factor H-binding protein profile completion Gonococcal Genetic Island profile completion Fron acquisition profile completion	ill be available in the filters section of the query interface. LOS alpha chain transferases profile completion LOS transport/export profile completion LIPI A biosynthesis: acyltransferases profile completion Lipid A biosynthesis: other profile completion Nuelestide excision repair profile completion Puties genes profile completion Protein glycosylation profile completion Prities genes profile completion Prine metabolism profile completion Private dehydrogenase completion RNA polymerase profile completion Small regulatory RNAs profile completion COMPORTING Completion COMPORTING Completion COMPORTING Completion COMPORTING Completion COMPORTING Completion COMPORTING Completion COMPORTING Completion COMPORTING Completion COMPORTING Completion UDP-GleNAc synthesis profile completion Wriff TASS profile completion eMLST (20 locus partial genes) profile completion eMLST (20 locus partial genes) profile completion PriF species profile completion Publications

A list of possible filters appears. Click any checkbox for a filter you would like to make available. Click 'Set options' when done. You can return to the default selection by clicking 'Default' followed by 'Set options'.

### 11.10.6 Modifying locus and scheme display options

Whether or not loci, schemes or scheme fields are displayed in result tables, isolate records, or within query dropdown boxes can all be set with default options when first defined. These attributes can, however, be overridden by a user, and these selections will be remembered between sessions.

The procedure to modify these attributes is the same for locus, schemes or scheme fields, so the steps for loci will be demonstrated only.

Click the appropriate link on the isolate contents page.

Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.
Search by combinations of loci (profiles)     Projects - main projects defined in database     Search by combinations of loci (profiles)     Projects - main projects defined in database.     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options for locus, schemes or scheme fields     Set display and query options
Breakdown       Export       Analysis       Miscellaneous         • Single field       • Export dataset       • Contigs       • Codon usage       • Description of database fields         • Unique combinations       • Scheme and alleles       • Sequences - XMFA / concatenated FASTA formats       • Codon usage       • Description of database fields         • Publications       • Sequence bin       • Sequence of the security of the secure secure security of the security of the security of t

Either select the locus id by querying for it directly.

Ouery:         Search   Browse   Profile/ST   List           Breakdown:         Isolate fields   Scheme/alleles   Publications           Links:         Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Т	Toggle: i
Query loci for Neisseria PubMLST database	
Please enter your search criteria below (or leave blank and submit to return all records). Matching loci will be returned and you will then be able to update their display and query settings.	
- Search criteria Display	
id v = v abcZ + i Order by: id v ascending v	
Display: 25	
Reset Submit	
1 record returned. Click the hyperlink for detailed information.	
Customize	
locus options	
id data type allele id format length length varies coding sequence orf genome position isolate display* main display* query field* analysis* abc2 DNA integer 433 false true 1 1176340 allele only false true true	
aUCz DVA integet 4-53 tabe uue i 1175340 aneteonin tabe uue uue * Default values are displayed for this field. These may be overridden by user preference.	

Designations can be queried using standard operators.

Alternatively, you can search by filtering loci by schemes. Click the 'Filter query by' header and select the scheme in the dropdown box.

								se submissions					
													Toggle: i
uer	y loci	for Neis	sseria	PubML	ST data	base	)						
Please	enter your	search criteria	below (or	leave blank a	nd submit to re	turn all re	ecords). Matchin	g loci will be returr	ned and you wi	ill then be ab	le to update the	eir display and query setting	js.
-Sea	arch criteria						— — Display-						
id		<b>→</b> =		•		+ [	Order b	y: id	•	ascending			
						_	Displa	iy: 25 👻 record	ds per page 👔	]			
	ilter query l	- 11					Action						
		data type:	•										
		id format:	-				Reset Subr	nit					
		th varies:	- ī										
		equence:	-										
		orf:	-										
	mate	h longest:	- -										
		pcr filter:	- i										
	р	robe filter:	- i										
		flag table:	- i										
	isolat	e display:		-									
	mai	n display:	<b>v</b> [	-									
	q	uery field:	<b>v</b> <u>i</u>										
		analysis:	<b>v</b> i										
		curator:		-	i								
		scheme: M	LST		•	i							
ecord	e returned	Click the hype	rlinke for a	detailed inform	ation								
		onex are hype	ining for c	ictanea morn	auon.								
Cus	tomize												
locus	options												
id	data type	allele id forma	t length l	enoth varies	coding sequer	ce orf o	enome position	isolate display*	main display*	query field*	analysis*		
abcZ	DNA	integer	433	false	true	1	1176340	allele only	false	true	true		
adk	DNA	integer	465	false	true	1	991951	allele only	false	true	true		
aroE	DNA	integer	490	false	true	2	2079469	allele only	false	true	true		
umC gdh	DNA DNA	integer integer	465 501	false false	true	1	1592943 1514394	allele only allele only	false false	true true	true		
dhC	DNA	integer	480	false	true	1	1453970	allele only	false	true	true		
		integer	450	false	true	1	965481	allele only	false	true	true		

Once loci have been selected, click Customize 'locus options'.

Cus	tomize											
locus	options											
id	data type	allele id forma	t length	length varies	coding sequenc	e orf g	enome position	isolate display*	main display*	query field*	analysis*	
abcZ	DNA	integer	433	false	true	1	1176340	allele only	false	true	true	
adk	DNA	integer	465	false	true	1	991951	allele only	false	true	true	
aroE	DNA	integer	490	false	true	2	2079469	allele only	false	true	true	
fumC	DNA	integer	465	false	true	1	1592943	allele only	false	true	true	
gdh	DNA	integer	501	false	true	1	1514394	allele only	false	true	true	
pdhC	DNA	integer	480	false	true	1	1453970	allele only	false	true	true	
pgm	DNA	integer	450	false	true	1	965481	allele only	false	true	true	

You can then choose to add or remove individual loci from the selection by clicking the appropriate checkboxes. At the bottom of the page are a number of attributes that you can change - clicking 'Change' will affect all selected loci.

Possible options for loci are:

• isolate\_display - Sets how the locus is displayed within an isolate record:

- allele only display only identifier
- sequence display the full sequence
- hide don't show at all
- main\_display Sets whether the locus is displayed in the main results table following a query.
- query\_field Sets whether the locus appears in dropdown list boxes to be used within queries.
- analysis Sets whether the locus can be used in data analysis functions.

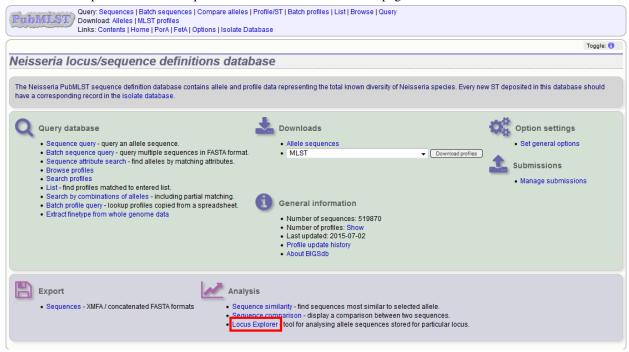
**Note:** Settings for loci can be overridden by those set for schemes that they are members of. For example, if you set a locus to be displayed within a main results table, but that locus is a member of a scheme and you set that scheme not to be displayed, then the locus will not be shown. Conversely, if you set a scheme to be displayed, but set its member locus not to be shown, then that locus will not be displayed (but other loci and scheme fields may be, depending on their independent settings).

# Data analysis plugins

# 12.1 Locus explorer

The locus explorer is a sequence definition database plugin. It can create schematics showing the polymorphic sites within a locus, calculate the GC content and generate aligned translated sequences.

Click 'Locus Explorer' from the sequence definition database contents page.



## 12.1.1 Polymorphic site analysis

Select the locus you would like to analyse in the Locus dropdown box. The page will reload.

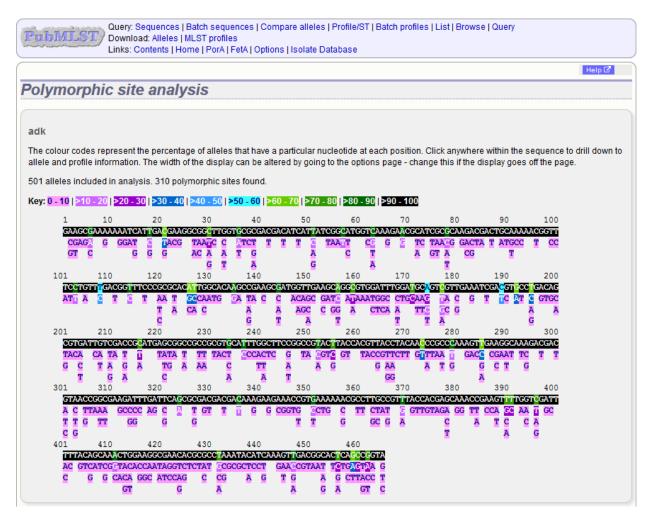
PubMLST / Downloa	Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse ad: Alleles   MLST profiles contents   Home   PorA   FetA   Options   Isolate Database	Query
		Help 🗗
Locus Explorer -	Neisseria locus/sequence definitions	
Please select locus for analy		
Locus adk	<ul> <li>Page will reload when changed</li> </ul>	
Select sequences		-Action-
1 All None	<ul> <li>Polymorphic Sites - Display polymorphic site frequencies and sequence schematic</li> <li>Codon - Calculate G+C content and codon usage</li> <li>Translate - Translate DNA to peptide sequences</li> </ul>	Submit

Select the alleles that you would like to include in the analysis. Variable length loci are limited to 2000 sequences or fewer since these need to be aligned. Select 'Polymorphic Sites' in the Analysis selection and click 'Submit'.

PubMLST Downlo	Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query bad: Alleles   MLST profiles Contents   Home   PorA   FetA   Options   Isolate Database	
		Help 🗗
Locus Explorer	- Neisseria locus/sequence definitions	
Please select locus for analy	lysis: ✔ Page will reload when changed	
Select sequences 496 497 498 499 500 501 T All None	Select analysis     Action     Polymorphic Sites     Display polymorphic site frequencies and sequence schematic     Submit     Codon - Calculate G+C content and codon usage     Translate - Translate DNA to peptide sequences	

If an alignment is necessary, the job will be submitted to the job queue and the analysis performed. If no alignment is necessary, then the analysis is shown immediately.

The first part of the page shows the schematic.



Clicking any of the sequence bases will calculate the exact frequencies of the different nucleotides at that position.

>40 - 50	>50 - 60   >	•60 - 70 <b>&gt;</b> 70	- 80 >80 - 90	>9
30	40	50	60	
GCGG <mark>C</mark> TTG	GT <mark>G</mark> CGCGAC	GACATCAT	TCGGCATGGTCA	AA(
		ТТ 🖸	TAATT CG	G
AC A A	TG	A	СТ	
GT	A	G	A	

te l	Explorer				Help 🛙
.~ 1				 	 
lk po	osition 51				
	eles included in anal	veie			
ase I T	Number of alleles P 401	ercentage of alle 80.04	eles MLST profiles 11130 / 11407		
			(97.57%)		
С	98	19.56	275 / 11407 (2.41%)		
A	1 (adk-351)	0.20	1 / 11407 (0.01%)		
G	1 (adk-413)	0.20	1 / 11407 (0.01%)		

The second part of the page shows a table listing nucleotide frequencies at each of the variable positions.

C (		410		42	0		20		0	45	•		c 0			1		A	G	
401	<b>B</b> ACIAC		omee		-	ACACGO	30	44		45			60	(* <b>11</b> 7)						
						GTCTCT														
C				GGC			C CG			TG		GC								
			GT			G	А			А	G			с						
			_				_			_		-	_							
cleotid	e fre	que	encie	s																
ition 🗢		0				lucleot		k o	br 77											
	A ¢			T ¢		%A ≑			%I≑	%-≑										
2	499	1	1	0	0		0.20		0.05											
3	498	0	2	1	0	99.40			0.20											
4	1	0	500	0	0	0.20		99.80												
5	0	500	1	0	0		99.80													
6	53	8	440	0	0		1.60													
9	499	0	2	0	0	99.60		0.40												
12	500	0	1	0	0			0.20												
13	500	0	1	0	0	99.80		0.20												
14	1	0	0	500	0	0.20			99.80											
15	0	488	1	12	0			0.20												
18	0	86	8	407	0			1.60												
21	0	315	1	185	0		62.87	0.20	36.93											
22	9	0	492	-	0	1.80		98.20												
23	500	1	0	0	0	99.80	0.20													
24	500	0	1	0	0	99.80		0.20												
27	3	488	0	10	0		97.41		2.00											
28	2	2	497	0	0		0.40													
29	2	0	499	0	0	0.40		99.60												
30	3	367	2	129	0	0.60	73.25	0.40	25.75											
31	0	11	0	490	0		2.20		97.80											
33	2	5	492	2	0	0.40	1.00	98.20	0.40											
	53	0	445	3	0	10.58		88.82	0.60											
36	55																			

#### See also:

- Investigating allele differences.
- Polymorphism analysis following isolate query.

### 12.1.2 Codon usage

Select the alleles that you would like to include in the analysis. Again, variable length loci are limited to 200 sequences or fewer since these need to be aligned. Click 'Codon'.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database
Locus Explorer - Neisseria locus/sequence definitions
Please select locus for analysis:
Locus: adk       Page will reload when changed         Select sequences       Analysis functions         459       Codon Calculate G+C content and codon usage         453       Translate DNA to peptide sequences         All None       None

The GC content of the alleles will be determined and a table of the codon frequencies displayed.

nt want in	LST/Dot	wnload: Allele	s   MLST profil	les	ipare alleles   Profile/ST   Batch profiles   List   Browse   Query is   Isolate Database
odon	Usage				
dk					
RF used: 1	1				
54 alleles	included in a	nalysis.			
C conten	it				
oding: GC st letter: G(					
	C 65.31% C 31.73%				
rd letter: G					
odons					
		sage of a giver en codon per '		ig its redunda	nt set (i.e. the set of codons which code for this codon's amino acid).
requency: I odon 🗢 🗛	Usage of giv	en codon per *	1000 codons. Frequency ¢	Number 🗢	nt set (i.e. the set of codons which code for this codon's amino acid).
requency: I odon 🔶 A GCA	Usage of giv Amino acid A	en codon per * Fraction ¢ 0.262	1000 codons. Frequency ¢ 17.353	Number ¢ 1248	nt set (i.e. the set of codons which code for this codon's amino acid).
odon 🔶 A GCA GCC	Usage of giv Amino acid A A A	en codon per 4 Fraction \$ 0.262 0.246	1000 codons. Frequency ¢ 17.353 16.254	Number ¢ 1248 1169	nt set (i.e. the set of codons which code for this codon's amino acid).
odon ¢ A GCA GCC GCC GCG	Usage of giv mino acid A A A A	en codon per * Fraction  O.262 0.246 0.389	1000 codons. Frequency ¢ 17.353 16.254 25.751	Number ♦ 1248 1169 1852	nt set (i.e. the set of codons which code for this codon's amino acid).
odon 🔶 A GCA GCC GCC GCC GCT	Usage of giv Amino acid 4 A A A A A	en codon per * Fraction ♦ 0.262 0.246 0.389 0.103	1000 codons. Frequency ♦ 17.353 16.254 25.751 6.813	Number ♦ 1248 1169 1852 490	nt set (i.e. the set of codons which code for this codon's amino acid).
odon ¢ A GCA GCC GCC GCC GCT TGC	Usage of giv Amino acid 4 A A A A A C	en codon per * Fraction \$ 0.262 0.246 0.389 0.103 0.987	1000 codons. Frequency ↓ 17.353 16.254 25.751 6.813 6.452	Number ♦ 1248 1169 1852 490 464	nt set (i.e. the set of codons which code for this codon's amino acid).
odon 🔶 A GCA GCC GCC GCC GCT	Usage of giv Amino acid 4 A A A A A	en codon per * Fraction ♦ 0.262 0.246 0.389 0.103	1000 codons. Frequency ♦ 17.353 16.254 25.751 6.813	Number ♦ 1248 1169 1852 490	nt set (i.e. the set of codons which code for this codon's amino acid).
equency: I odon \$ A GCA GCC GCC GCT TGC TGT	Usage of giv mino acid 4 A A A A C C C	en codon per * Fraction ♦ 0.262 0.246 0.389 0.103 0.987 0.013	1000 codons. Frequency ↓ 17.353 16.254 25.751 6.813 6.452 0.083	Number 1248 1169 1852 490 464 6	nt set (i.e. the set of codons which code for this codon's amino acid).
equency: I odon + A GCA GCC GCC GCT TGC TGC TGT GAC	Usage of giv mino acid A A A A C C C D	en codon per Fraction ¢ 0.262 0.246 0.389 0.103 0.987 0.013 0.747	1000 codons. Frequency ◆ 17.353 16.254 25.751 6.813 6.452 0.083 91.073	Number ♦ 1248 1169 1852 490 464 6 6550	nt set (i.e. the set of codons which code for this codon's amino acid).
odon  A GCA GCC GCG GCT TGC TGT GAC GAT	Usage of giv mino acid ¢ A A A C C C D D	en codon per 3 Fraction ¢ 0.262 0.246 0.389 0.103 0.987 0.013 0.747 0.253	1000 codons. Frequency ♦ 17.353 16.254 25.751 6.813 6.452 0.083 91.073 30.812	Number ↓ 1248 1169 1852 490 464 6 6550 2216	nt set (i.e. the set of codons which code for this codon's amino acid).
equency: 1 odon  A GCA GCC GCG GCT TGC TGT GAC GAT GAA	Usage of giv mino acid ¢ A A A A C C D D E	en codon per * Fraction ♦ 0.262 0.246 0.389 0.103 0.987 0.013 0.747 0.253 0.916	1000 codons. Frequency ♦ 17.353 16.254 25.751 6.813 6.452 0.083 91.073 30.812 82.397	Number ♦ 1248 1169 1852 490 464 6 6550 2216 5926	nt set (i.e. the set of codons which code for this codon's amino acid).
equency: I GCA GCC GCC GCG GCT TGC TGC TGT GAC GAT GAA GAG	Usage of giv mino acid 4 A A A C C D D E E	en codon per * Fraction ↓ 0.262 0.246 0.389 0.103 0.987 0.013 0.747 0.253 0.916 0.084	1000 codons. Frequency ♦ 17.353 16.254 25.751 6.813 6.452 0.083 91.073 30.812 82.397 7.564	Number ♦ 1248 1169 1852 490 464 6 6550 2216 5926 544	nt set (i.e. the set of codons which code for this codon's amino acid).
equency: 1 GCA GCC GCC GCT TGC TGT GAC GAT GAA GAA GAG TTC	Usage of giv mino acid 4 A A A C C D D D E E F	en codon per * Fraction ↓ 0.262 0.246 0.389 0.103 0.987 0.013 0.747 0.253 0.916 0.084 0.594	1000 codons. Frequency ♦ 17.353 16.254 25.751 6.813 6.452 0.083 91.073 30.812 82.397 7.564 15.295	Number ◆ 1248 1169 1852 490 464 6 5550 2216 5926 544 1100	nt set (i.e. the set of codons which code for this codon's amino acid).
odon \$ A       GCA       GCC       GAC       GAC       GAC       TTC       TTT	Usage of giv mino acid A A A C C C D D E E F F	en codon per * Fraction  Fraction  Fraction  Outline O	1000 codons. Frequency ◆ 17.353 16.254 25.751 6.813 6.452 0.083 91.073 30.812 82.397 7.564 15.295 10.470	Number ♦ 1248 1169 1852 490 464 6 6550 2216 5926 544 1100 753	nt set (i.e. the set of codons which code for this codon's amino acid).
equency: 1 odon \$ A GCA GCC GCG GCT TGC TGT GAC GAT GAA GAA GAA GAA GAA GAA GAA	Usage of giv mino acid ¢ A A A C C C D D E E F F G	en codon per * Fraction   Fraction   Output  Output  Fraction   Output   1000 codons. Frequency ◆ 17.353 16.254 25.751 6.813 6.452 0.083 91.073 30.812 82.397 7.564 15.295 10.470 0.542	Number 1248 1169 1852 490 464 6 6550 2216 5926 544 1100 753 39	nt set (i.e. the set of codons which code for this codon's amino acid).	
requency: 1       odon \$ A       GCA       GCC       GCG       GCT       TGT       GAC       GAT       GAA       GAG       TTC       TTT       GGA       GGC	Usage of giv mino acid ( A A A A C C D D C D E E F F G G G	en codon per 1 Fraction ¢ 0.262 0.246 0.389 0.103 0.987 0.013 0.747 0.253 0.916 0.084 0.594 0.406 0.007 0.765	1000 codons. Frequency ♦ 17.353 16.254 25.751 6.813 6.452 0.083 91.073 30.812 82.397 7.564 15.295 10.470 0.542 59.497	Number ♦ 1248 1169 1852 490 464 6 6550 2216 5926 544 1100 753 39 4279	nt set (i.e. the set of codons which code for this codon's amino acid).
equency: 1 odon ♦ A GCA GCC GCC GCT TGC TGC TGC GAC GAA GAA GAA TTC TTT GAA GAG GGC GGC GGC	Usage of giv unino acid A A A A C C C C D D D E E F F G G G	en codon per *	1000 codons. Frequency ◆ 17.353 16.254 25.751 6.813 6.452 0.083 91.073 30.812 82.397 7.564 15.295 10.470 0.542 59.497 0.042	Number ♦ 1248 1169 1852 490 464 6 550 2216 5926 544 1100 753 39 4279 3	nt set (i.e. the set of codons which code for this codon's amino acid).
equency: 1 odon \$ A GCA GCC GCC GCT TGC TGC TGT GAC GAC GAA TTC TTT GGA GGG GGG GGG	Usage of giv unino acid 4 A A A A A C C C D D D D D D E E E F F G G G G G G	en codon per *	1000 codons. Frequency ♦ 17.353 16.254 25.751 6.813 91.073 30.812 82.397 7.564 15.295 10.470 0.542 59.497 0.042 17.559	Number ♦ 1248 11652 1852 490 464 6 6550 2216 5926 544 1100 753 39 4279 3 1270	nt set (i.e. the set of codons which code for this codon's amino acid).

## 12.1.3 Aligned translations

If a DNA coding sequence locus is selected, an aligned translation can be produced.

Select the alleles that you would like to include in the analysis. Again, variable length loci are limited to 200 sequences or fewer since these need to be aligned. Click 'Translate'.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database
Locus Explorer - Neisseria locus/sequence definitions
Please select locus for analysis:
Locus: adk Page will reload when changed Select sequences Analysis functions Polymorphic sites Display polymorphic site frequencies and sequence schematic Codon Calculate G+C content and codon usage Codon Calculate G+C content and codon usage Translate Translate DNA to peptide sequences

An aligned amino acid sequence will be displayed.

33_1 -	·········
34_1	
5_1	Т.
6_1	
7_1	
8_1	
9_1	
0_1	К.
1_1	
2_1	
3_1	
4_1	
5_1	
6_1	
7_1	
3_1	
9_1	
1	
1_1	
2_1	КТ.
3_1	
1	IA
5_1	
6_1	
7 1	G
8_1	
9_1	
D_1	s
1_1	
2_1	
3 1	
4 1	I
	eakkiideggivrddiiiigwykeriaqddcwngflfdgfprtlaqaeawyeagvdldavyeidypdsvivdrwsgrrvhlasgrtyhytynppkyegkdd
nsensus	eakkiideGelvRddiiiGRVKekiaQdDCkNGFlFDGFPRTLAQAEAMVEAGVDLDAVVEIDVPDSVIVDRMSGRRVHLASGRTHVIYNPPRVEGRDD
	110 120 130 140 150
L	
1	
1	
1 1 1	
1	
1	
1	

If there appear to be a lot of stop codons in the translation, it is possible that the orf value in the *locus definition* is not set correctly.

## 12.2 Field breakdown

The field breakdown plugin for isolate databases displays the frequency of each value for fields stored in the isolates table. *Allele and scheme field breakdowns* are handled by a different plugin.

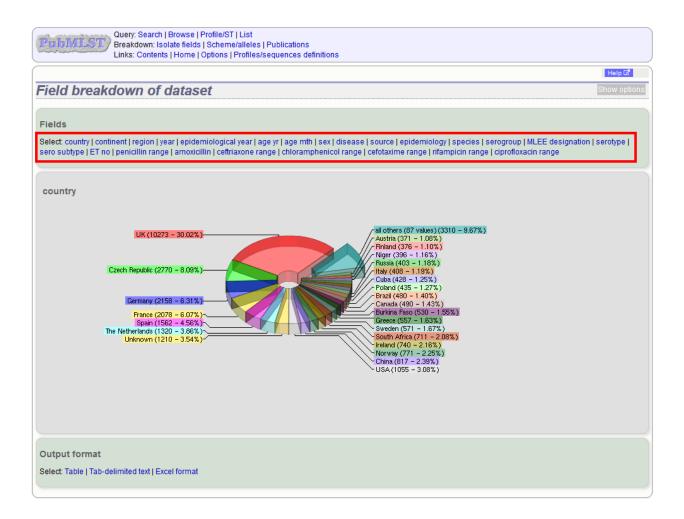
The breakdown function can be selected for the whole database by clicking the 'Single field' link in the Breakdown section of the main contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diver	
	<ul> <li>- including isolate table field handling.</li> <li>- y options for locus, schemes or scheme fields.</li> <li>Isolate: 34221</li> <li>- Last updated: 2015-07-02</li> <li>- Update history</li> <li>- About BIGSdb</li> </ul>
Iwo field     Contigs     Pres	

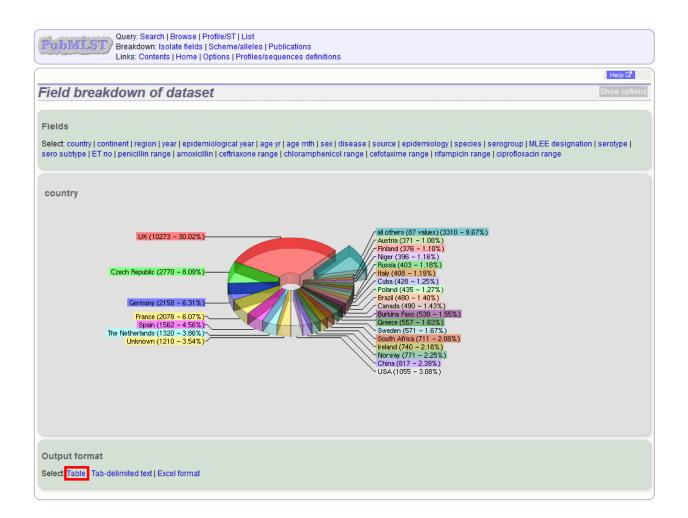
Alternatively, a breakdown can be displayed of the dataset returned from a query by clicking the 'Fields' button in the Breakdown list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

5         M00282207         UK         2000         invasive (unspecified/other)         Neisseria meningitidis         W         1101         ST-22 complex         Image: Complex <thimage: complex<="" th="">         Image: Complex<th>5 1</th><th></th><th></th><th></th><th></th><th>· · · ·</th><th>-</th><th></th><th></th><th></th><th></th><th></th><th></th></thimage:>	5 1					· · · ·	-						
7891       B54; Z1054       Finland       1975       invasive (unspecified/other)       Neisseria meningitidis       A       5       ST-5 complex/subgroup III       20       9       F3-1         M00242007       UK       2000       invasive (unspecified/other)       Neisseria meningitidis       B       1102       ST-16 complex       14       14         0021/84       Czech Republic 1984       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup /II       18-1       3       F5-1         129       B92, Z1092       Germany       1964       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup /II       18-1       3       F5-1         0909/89       Czech Republic 1989       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup /II       5-2       10       F5-1         3       139M       B99, Z1099       Philippines       1995       invasive (unspecified/other)       Neisseria meningitidis       X       117       14       14       14       14       14       14       14       14       14       14       14       14       16       14       16       16       16											7	16	
B         M00242007         UK         2000         invasive (unspecified/other)         Neisseria meningitidis         B         1102         ST-18 complex         14           0         0021/84         Czech Republic         1984         invasive (unspecified/other)         Neisseria meningitidis         W         114         ST-22 complex         14         Image: Circle Action (Circle Action (C	6 1	100282207							1101	ST-22 complex			
9         0021/84         Czech Republic 1984         invasive (unspecified/other)         Neisseria meningitidis         W         114         ST-22 complex         C         C           0         6748         B73; Z1073         Canada         1971         invasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup I/II         18-1         3         F5-1           1         129         B92; Z1092         Germany         1964         invasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup I/II         5-2         10         F5-1           2         0090/89         Czech Republic 1985         invasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup I/II         5-2         10         F5-1           4         012095         Czech Republic 1995         invasive (unspecified/other)         Neisseria meningitidis         X         117         14         5-2         10         F5-1           5         1         Germany 1999         carrier         Neisseria meningitidis         B         854         ST-18 complex         5-2         10         F5-1           6         2         Germany 1999         carri	7		B54; Z1054							ST-5 complex/subgroup III	20		F3-1
0         6748         B73;Z1073         Canada         1971         invasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup //II         18-1         3         F5-1           1         129         B92;Z1092         Germany         1994         invasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup //II         5-2         100         F3-6           3         139M         B99;Z1099         Philippines         1986         Neisseria meningitidis         A         1         ST-1 complex/subgroup //II         5-2         100         F5-1           4         012095         Czech Republic 1995         invasive (unspecified/other)         Neisseria meningitidis         X         117         14         14           5         1         Germany         1999         carrier         Neisseria meningitidis         B         854         ST-18 complex         14         14           6         2         Germany         1999         carrier         Neisseria meningitidis         B         19         ST-18 complex         16         16         16         14         16         14         16         16         16         16         16 <td>8 1</td> <td>100242007</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>1102</td> <td>ST-18 complex</td> <td></td> <td>14</td> <td></td>	8 1	100242007							1102	ST-18 complex		14	
1       129       B92, Z1092       Germany       1964       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup //II       5-2       10       F3-6         2       0090/99       Czech Republic       1989       invasive (unspecified/other)       Neisseria meningitidis       B       1015       ST-1 complex/subgroup //II       5-2       10       F3-6         3       139M       B99, 2109       Philippines       1988       Neisseria meningitidis       X       117       5-2       10       F5-1         4       012095       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       X       117       5-2       10       F5-1         5       1       Germany       1999       carrier       Neisseria meningitidis       X       117       5-2       10       F5-1         6       2       Germany       1999       carrier       Neisseria meningitidis       W       174       ST-18 complex	9	0021/84		Czech Republic	1984	invasive (unspecified/other)	Neisseria meningiti	idis W	114	ST-22 complex			
20090/89Czech Republic1989invasive (unspecified/other)Neisseria meningitidisB1015ST-32 complex/ET-5 complex7163139MB99; 2109Philippines1988Neisseria meningitidisA1ST-1 complex/subgroup //II5-210F5-14012095Czech Republic1995invasive (unspecified/other)Neisseria meningitidisX117T1451Germany1999carrierNeisseria meningitidisE864141462Germany1999carrierNeisseria meningitidisB854ST-18 complex1673Germany1999carrierNeisseria meningitidisB19ST-18 complex1684Germany1999carrierNeisseria meningitidisB19ST-18 complex169S131B213; Z1213Ghana1973invasive (unspecified/other)Neisseria meningitidisNG198ST-180 complex1616Germany1999carrierNeisseria meningitidisNG198ST-180 complex1616Germany1999carrierNeisseria meningitidisNG198ST-180 complex1627Germany1999carrierNeisseria meningitidisNG198ST-180 complex1627Germany1999carrierNeisseria meningitidisNG19	10	6748	B73; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningiti	idis A	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
3       139M       B99; Z1099       Philippines       1968       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/II       5-2       10       F5-1         4       012095       Germany       1999       carrier       Neisseria meningitidis       X       117       14       14       14         6       2       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       14       14         7       3       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       16       16         9       S3131       B213; Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis       A       4       ST-4 complex       16       16         0       5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-180 complex       17       13-1       F1-5         0       5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       16       16       Germany       1999       carrier       Neisseria meningitidis       NG       <	11	129	B92; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningiti	idis A	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
4       0120/95       Czech Republic 1995       invasive (unspecified/other)       Neisseria meningitidis       X       117       14         5       1       Germany       1999       carrier       Neisseria meningitidis       E       864       14       14         6       2       Germany       1999       carrier       Neisseria meningitidis       E       864       14       14         7       3       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       14       14         9       S3131       B213, Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis       B       19       ST-18 complex       14       14         0       5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       14       14         0       5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       17       13-1       F1-5         1       6       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       14       <	12	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningiti	idis B	1015	ST-32 complex/ET-5 complex	7	16	
5       1       Germany       1999       carrier       Neisseria meningitidis       E       864       0       0       0         6       2       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       0       0         7       3       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       0       0         8       4       Germany       1999       carrier       Neisseria meningitidis       B       19       ST-18 complex       0       0         9       S3131       B213,Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis       A       4       ST-48 complex       0       0         0       5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       0       0         1       6       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       0       0       0       0       0       0       0       0       0       0       0       0       0       0	13	139M	B99; Z1099	Philippines	1968		Neisseria meningiti	idis A	1	ST-1 complex/subgroup I/II	5-2		F5-1
6       2       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       Image: Compl	14	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningiti	idis X	117			14	
7       3       Germany       1999       carrier       Neisseria meningilidis       W       174       ST-174 complex       ST-174 comp	15	1		Germany	1999	carrier	Neisseria meningiti	idis E	864				
8       4       Germany       1999       carrier       Neisseria meningitidis       B       19       ST-18 complex       Image: ST-18	16	2		Germany	1999	carrier	Neisseria meningiti	idis B	854	ST-18 complex			
9       S3131       B213;Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis       A       4       ST-4 complex/subgroup IV       7       13-1       F1-5         0       5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -       -	17	3		Germany	1999	carrier	Neisseria meningiti	idis W	174	ST-174 complex			
0       5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       Image: Com	18	4		Germany	1999	carrier	Neisseria meningiti	idis B	19	ST-18 complex			
1       6       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex         2       7       Germany       1999       carrier       Neisseria meningitidis       E       60       ST-60 complex         3       8       Germany       1999       carrier       Neisseria meningitidis       B       32       ST-32 complex/EF-5 complex         4       S4355       B227;Z127       Denmark       1974       invasive (unspecified/other)       Neisseria meningitidis       B       930       ST-334 complex       9       F3-1         5       9       Germany       1999       carrier       Neisseria meningitidis       B       930       ST-334 complex       9       F3-1	19	S3131	B213; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningiti	idis A	4	ST-4 complex/subgroup IV	7	13-1	F1-5
2     7     Germany     1999     carrier     Neisseria meningitidis     E     60     ST-60 complex       3     8     Germany     1999     carrier     Neisseria meningitidis     B     32     ST-32 complex/ET-5 complex       4     S4355     B227; Z1227     Denmark     1974     invasive (unspecified/other)     Neisseria meningitidis     A     5     ST-5 complex/subgroup III     5-1     9     F3-1       5     9     Germany     1999     carrier     Neisseria meningitidis     B     930     ST-334 complex	20	5		Germany	1999	carrier	Neisseria meningiti	idis NG	198	ST-198 complex			
3     8     Germany     1999     carrier     Neisseria meningitidis     B     32     ST-32     complex/ET-6     complex	21	6		Germany	1999	carrier	Neisseria meningiti	idis NG	198	ST-198 complex			
4 S4355 B227;Z1227 Denmark 1974 invasive (unspecified/other) Neisseria meningitidis A 5 ST-5 complex/subgroup III 5-1 9 F3-1 5 9 Germany 1999 carrier Neisseria meningitidis B 930 ST-334 complex	22	7		Germany	1999	carrier	Neisseria meningiti	idis E	60	ST-60 complex			
5 9 Germany 1999 carrier Neisseria meningitidis B 930 ST-334 complex	23	8		Germany	1999	carrier	Neisseria meningiti	idis B	32	ST-32 complex/ET-5 complex			
	24	S4355	B227; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningiti	idis A	5	ST-5 complex/subgroup III	5-1	9	F3-1
nalysis tools:	25	9		Germany	1999	carrier	Neisseria meningiti	idis B	930	ST-334 complex			
eakdown: Fields Two Field Codons Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status	a	8 S4355 9 lysis tools	:	Germany Denmark Germany	1999 1974 1999	carrier invasive (unspecified/other) carrier	Neisseria meningiti Neisseria meningiti Neisseria meningiti	idis B idis A idis B	32 5 930	ST-32 complex/ET-5 complex ST-5 complex/subgroup III ST-334 complex	5-1	9	F3-1
		nalysis: 🛛 BL	JRST J Prese	noe/Absence	Senome	Comparator BLAST							

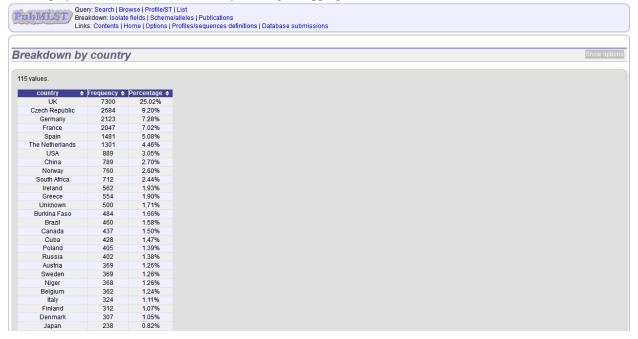
A series of charts will be displayed. Pick the field to display from the list at the top.

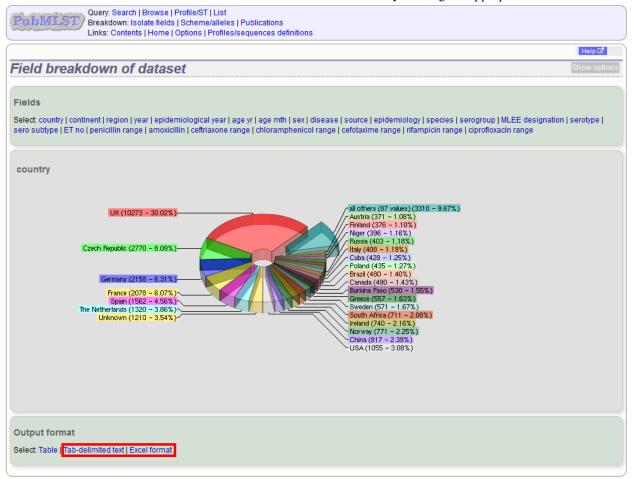


The values used to generate the chart can be displayed or extracted by clicking the 'Table' link at the bottom of the page.



This displays a table that can be ordered by clicking the appropriate header.





The data can also be downloaded in tab-delimited text or Excel formats by clicking the appropriate links.

# 12.3 Two field breakdown

The two field breakdown plugin displays a table breaking down one field against another, e.g. breakdown of serogroup by year.

The analysis can be selected for the whole database by clicking the 'Two field breakdown' link on the main contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic pro database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that	
<ul> <li>Query database</li> <li>Search database - advanced queries.</li> <li>Browse database - peruse all records.</li> <li>Search by combinations of loci (profiles) - including partial matching.</li> <li>List query - find isolates by matching a field to an entered list.</li> <li>Projects - main projects defined in database.</li> </ul>	Ids. Submissions • Manage submissions • General information • Isolates: 34221 • Last updated: 2015-07-02 • Update history • About BIGSdb
	ellaneous cription of database fields

Alternatively, a two field breakdown can be displayed of the dataset returned from a query by clicking the 'Two field' button in the Breakdown list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

ſ	onanig	on the		une	autuouse.							
6			UK		invasive (unspecified/other)			1101	S1-22 complex			
7	7891	B54; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitid	is A	5	ST-5 complex/subgroup III	20	9	F3-1
8			UK	2000	invasive (unspecified/other)	Neisseria meningitid	is B	1102	ST-18 complex		14	
9	0021/84		Czech Republic	1984	invasive (unspecified/other)	Neisseria meningitid	is W	114	ST-22 complex			
10	6748	B73; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitid	is A	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
11	1 129	B92; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitid	is A	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
12	2 0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningitid	is B	1015	ST-32 complex/ET-5 complex	7	16	
13	3 139M	B99; Z1099	Philippines	1968		Neisseria meningitid	is A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
14	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitid	is X	117			14	
15	5 1		Germany	1999	carrier	Neisseria meningitid	is E	864				
16	6 2		Germany	1999	carrier	Neisseria meningitid	is B	854	ST-18 complex			
17	7 3		Germany	1999	carrier	Neisseria meningitid	is W	174	ST-174 complex			
18	3 4		Germany	1999	carrier	Neisseria meningitid	is B	19	ST-18 complex			
19	S3131	B213; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitid	is A	4	ST-4 complex/subgroup IV	7	13-1	F1-5
20	) 5		Germany	1999	carrier	Neisseria meningitid	is NG	198	ST-198 complex			
21	1 6		Germany	1999	carrier	Neisseria meningitid	is NG	198	ST-198 complex			
22	2 7		Germany	1999	carrier	Neisseria meningitid	is E	60	ST-60 complex			
23	3 8		Germany	1999	carrier	Neisseria meningitid	is B	32	ST-32 complex/ET-5 complex			
24	\$4355	B227; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitid	is A	5	ST-5 complex/subgroup III	5-1	9	F3-1
25	5 9		Germany	1999	carrier	Neisseria meningitid	is B	930	ST-334 complex			
	_	elds Two Fi	ence/Absence	Genom	norphic sites Combinations	Schemes/alleles F	Publications	Seque	nce bin Tag status			

Select the two fields you wish to breakdown and how you would like the values displayed (percentage/absolute values and totaling options).

Query: Search   Browse   Profile/ST Breakdown: Isolate fields   Scheme Links: Contents   Home   Options   F	/alleles   Publications	Database submissions		
Two field breakdown of dataset	t			Show options
Here you can create a table breaking down one field by a Select fields Field 1: clonal complex (MLST) Field 2: serogroup	nother, e.g. breakdown of sero — Display — — — — @ values only values and percentages percentages only	Calculate percentages by-	Action Reset Submit	

Click submit. The breakdown will be displayed as a table. Bar charts will also be displayed provided the number of returned values for both fields are fewer than 30.

wo field breakdown of dataset       Image: Complex (MLST) by serogroup:         streakdown of clonal_complex (MLST) by serogroup:       Streakdown of clonal_complex (MLST) by serogroup:         Aver get by values only:       Image: Complex (MLST) by serogroup:         Aver get by values and prombages       Streakdown of clonal_complex (MLST) by serogroup:         Clonal_complex (MLST) (Values and prombages)       Streakdown of clonal_complex (MLST) (Values and prombages)         Streakdown of clonal_complex (MLST) (Values and prombages)       Streakdown of clonal_complex (MLST) (Values and prombages)         Streakdown of clonal_complex (MLST) (Values and prombages)       Streakdown of clonal_complex (Values and prombages)         Streakdown of clonal_complex (MLST) (Values and prombages)       Streakdown of clonal_complex (Values and prombages)         Streakdown of clonal_complex (MLST) (Values and prombages)       Streakdown of clonal_complex (Values and prombages)         Streakdown of clonal_complex (Values and prombages)       Streakdown of clonal_complex (Values and prombages)         Streakdown of clonal_complex (Values and prombages)       Streakdown of clonal_complex (Values and prombages)         Streakdown of clonal_complex (Values and prombages)       Streakdown of clonal_complex (Values and prombages)         Streakdown of clonal_complex (Values and prombages)       Streakdown of clonal_complex (Values and prombages)         Streakdown of clonal_complex (Values and prombages)       Streakdown of clonal_complex (	Irreakdown of clonal_complex (MLST) by serogroup:         elected options: Display values only:         Ares       Show         Reverse       Values and percentages         clonal_complex (MLST)       I       Serogroup         clonal_complex (MLST)       I       Serogroup         clonal_complex (MLST)       I       Serogroup         clonal_complex (MLST)       I       Serogroup         clonal_complex (MLST)       I       I       I         ST-1 complex/subgroup III       I       I       I         ST-1 complex/subgroup III       I       I       I         ST-13C complex       I       I       I       I         ST-14 complex/subgroup III       I       I       I       I         ST-18 complex       I       I       I       I       I         ST-22 complex       I       I       I       I       I       I         ST-22 complex       I       I       I       I       I       I       I         ST-22 complex       I       I       I       I       I       I       I       I         ST-24 complex/subgroup II       I       I       I       I       I										He		
Heded options: Display values only.         Axes       Show         Reverse       Values and percentages         Ional complex (MLST) ◆ A ◆ B ◆ C ◆ E ◆ N6 ◆ W ◆ K ◆ K ◆ K ◆ K ◆ K ◆ K ◆ K ◆ K ◆ K	Accord options: Display values only:         Accord options: Display values only:         Reverse       Values and percentages         tomai complex (till S1)       A       B       C       B       C       Total       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C <td< th=""><th>vo field breakdo</th><th>wn</th><th>of d</th><th>lata</th><th>set</th><th></th><th></th><th></th><th></th><th>Show</th></td<>	vo field breakdo	wn	of d	lata	set					Show		
Beted options: Display values only.         Reverse       Values and percentages         Serogroup         Serogroup         No value       1       A         Serogroup	Aves       Show         Reverse       Values and percentages         Jonato_complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-102 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-102 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-102 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-102 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-102 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-102 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-102 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-102 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-22 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-22 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-22 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-22 complex (full_ST)       Image: Complex (full_ST)       Image: Complex (full_ST)         ST-22 complex (full_ST)       Image: Complex (full_ST)												
Show         Reverse       Serogroup         Values and percentages         Serogroup         Serogroup         No value       1       1       1         Serogroup         1       2       3         Serogroup       1       1         Serogroup       5         ST-102 complex       1       1       1         ST-226 complex       1       1         ST-248 complex       1       1         ST-248 complex       1       1         ST-248 complex       1       1         ST-32 complex/ST-2 complex <th colspan<="" th=""><th>alected options: Display values only.         Aver set         values and percentages             complex (ML ST)       Image: Complex (ML ST)             ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ST)       Image: Complex (ST)         ST-for complex (ST)       Image: Complex (ST)     &lt;</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></th>	<th>alected options: Display values only.         Aver set         values and percentages             complex (ML ST)       Image: Complex (ML ST)             ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ST)       Image: Complex (ST)         ST-for complex (ST)       Image: Complex (ST)     &lt;</th> <th></th>	alected options: Display values only.         Aver set         values and percentages             complex (ML ST)       Image: Complex (ML ST)             ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ML ST)       Image: Complex (ML ST)         ST-for complex (ST)       Image: Complex (ST)         ST-for complex (ST)       Image: Complex (ST)     <											
Aves       Show         Reverse       Values and percentages         Strate complex (MLST) + A + B + C + C + K + W + X + V + Z + Total + Novalue       Novalue       1       3       1       4       1       1 12         Strate complex (MLST) + A + B + C + C + K + W + X + V + Z + Total + Novalue       Novalue       1       3       1       4       1       1 12         Strate complex (MLST) + C + K + W + X + V + Z + Total + Novalue       Novalue       1       1       1 13         Strate complex (MLST) + A + B + C + K + W + X + V + Z + Total + Novalue       Novalue       1       1       1 13         Strate complex (MLST) + A + B + K + W + X + V + Z + Total + Novalue       Novalue       1       1       13         Strate complex (MLST) + A + B + K + W + X + V + Z + Total + Novalue       Novalue       1       1       13         Strate complex (MLST) + A + B + K + W + X + V + Z + Total + Novalue       Novalue       1       1       13         Strate complex (MLST) + A + B + K + W + X + V + X + Y + X + Y + X + X + Y + X + X + X	Aves         Show           Reverse         Values and percentages           Stand         Stand percentages           Stand         Stand percentages           Stand percentages         Stand percentages				51) Dy	sero	grou	p:					
Values and percentages         Stropplex (ML S1) \$        A \$        B \$        C \$        Stropplex (ML S1) \$        A \$        B \$        C \$        C onal \$ <thc \$="" <="" onal="" tr="">        ST-14 comp</thc>	eventses         Seventses         Seventses         No value       1         No value       1         ST - Complex       5         ST - Complex       5         ST - Complex       5         ST - 1       2         ST - 1       2         ST - 1       2         ST - 1       2         ST - 1       2       2         ST - 1       2       2         ST - 20       Complex       1         ST - 20       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2       2 <th 2"2"2<<="" colspan="2" th=""><th>elected options: Display value</th><th>s only.</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></th>	<th>elected options: Display value</th> <th>s only.</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>		elected options: Display value	s only.								
Seregroup           clonal_complex (MLST) ◆ A ◆ B ◆ C ◆ E ◆ NG ◆ N ◆ X ◆ Y ◆ Z ◆ Total ◆           No value         1         3         1         4         1         1         12           ST-domplex/Subgroup III         13         1         4         1         1         12           ST-domplex/Subgroup III         13         1         4         1         1         12           ST-domplex/Subgroup III         13         1         4         1         1         12           ST-162 complex         5         -         -         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         5         1         1         1         5         5         5         5         5         5         5         5         5         5	Sergroup         Sergroup         St-1 complex (MLS1)       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A       A	Axes — Show —											
Seregroup           clonal_complex (MLST) ◆ A + B + C + E + NG + W ◆ X + Y + Z + Total + No value           No value         1         1         1         1         1           ST-1 complex/Subgroup /// 13         3         1         1         4         1         1         12           ST-1 complex/Subgroup /// 13         1         4         1         1         11         12           ST-162 complex         5         -         -         5         5         -         5           ST-167 complex         5         -         1         2         3         3           ST-178 complex         1         1         1         3         3         -         3           ST-186 complex         2         2         2         2         2         2           ST-226 complex         1         1         1         1         1         3           ST-32 complex/Cluster A3         -         1         1         1         1           ST-32 complex         2         1         3         1         1         1           ST-32 complex         1         1         1         1         1         1	Sergroup           Sergroup           Sergroup           ST-10 complex ML S1 3         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1 <th colsp<="" td=""><td>Reverse Values and</td><td>perce</td><td>entages</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th>	<td>Reverse Values and</td> <td>perce</td> <td>entages</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	Reverse Values and	perce	entages								
clonal_complex (MLST)       A       A       B       C       E       W A       X       Y       Z       Total 4         No value       1       3       1       1       4       1       1       12         ST-1 complex/Subgroup III       13       -       1       1       12       13         ST-110 complex/Subgroup IIII       13       -       16       5       5       5         ST-162 complex       5       -       1       2       3       5       5       1       1       12         ST-174 complex       -       1       1       1       2       3       3       5       5       1       1       1       3         ST-178 complex       -       1       1       1       1       1       1       3       3       5       1       1       1       1       3       3       3       5       1       1       1       1       1       3       3       3       3       5       3       1       1       1       1       3       3       3       3       3       3       3       3       3       3       3       3	Clone Complex (ML ST)       A       B       B       C       C       F       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C       C												
Novalue         1         3         1         1         4         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1 </td <td>No value         1         3         1         4         1         1         1           ST-1 complex/subgroup (II         13         1         1         1         13         1         1         14         1         11         12           ST-1 complex/subgroup (II         13         1         1         1         13         15         1         14         15           ST-162 complex         5         1         1         2         3         3         15         1         2         3           ST-174 complex         1         1         1         1         3         3         1         1         3           ST-18 complex         3         2         2         2         2         3         1         1         1           ST-28 complex         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>A 77-4-1</td> <td></td>	No value         1         3         1         4         1         1         1           ST-1 complex/subgroup (II         13         1         1         1         13         1         1         14         1         11         12           ST-1 complex/subgroup (II         13         1         1         1         13         15         1         14         15           ST-162 complex         5         1         1         2         3         3         15         1         2         3           ST-174 complex         1         1         1         1         3         3         1         1         3           ST-18 complex         3         2         2         2         2         3         1         1         1           ST-28 complex         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1         1									A 77-4-1			
ST-1 complex/ET-37 complex     16     1     16     16       ST-162 complex     16     1     16       ST-162 complex     1     1     1     16       ST-167 complex     1     1     1     1       ST-167 complex     1     1     1     1       ST-174 complex     1     1     1     1       ST-186 complex     3     1     1     1       ST-198 complex     3     2     2       ST-22 complex     1     2     3       ST-22 complex     1     1     1       ST-256 complex     1     1     1       ST-256 complex     6     1     1       ST-41 complex/Lineage 3     5     1     1       ST-42 complex/Subgroup IV     8     1     1       ST-46 complex/Subgroup IV     8     1     1       ST-50 complex     1     1     1       ST-50 complex/Subgroup III     6	ST-1 complex/E-73 complex       16       13         T-11 complex/E-73 complex       16       16         ST-162 complex       5       1       2         ST-173 complex       1       1       1         ST-174 complex       1       1       1         ST-174 complex       1       1       1         ST-178 complex       2       2         ST-22 complex       2       2         ST-22 complex       1       1         ST-23 complex/E-3       1       1         ST-24 complex       1       1         ST-32 complex/E-5 complex       2       1         ST-34 complex/E-4       1       1         ST-34 complex/E-5       1       1         ST-54 complex/E-5       1       1         ST-40 complex/E-15       1       1         ST-40 complex/E-15       2       2         ST-40 complex/E-15       2       2         ST-40 complex/E-15       1       1         ST-40 complex/E-15       2       2         ST-50 complex/E-107       1       1         Total       2       2       2         ST-50 complex/E-107       1						vv ≑						
Ti-11 complex/ET-37 complex       16       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       III       IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	Ti-Ti complex     16     1     2       ST-162 complex     5     1     2       ST-176 complex     1     1     1       ST-176 complex     1     1     1       ST-176 complex     1     1     1       ST-176 complex     3     1     1       ST-176 complex     3     1     1       ST-18 complex     3     2     2       ST-22 complex     2     2       ST-22 complex     1     1       ST-32 complex     1     1       ST-32 complex     1     1       ST-44 complex/subgroup IV     8     1       ST-45 complex     1     1       ST-45 complex     1     1       ST-4144 complex/subgroup IV     8       ST-5 complex/subgroup IV     6       ST-5 complex/subgroup IV     1       Group IX     1 <td></td> <td></td> <td>5 1</td> <td></td> <td>4</td> <td></td> <td></td> <td></td> <td></td> <td></td>			5 1		4							
ST-162 complex       5       0       0       0       5         ST-174 complex       1       1       1       2       3         ST-174 complex       1       1       1       3         ST-178 complex       3       1       1       1       3         ST-178 complex       3       1       1       1       3         ST-180 complex       3       2       2       3         ST-190 complex       3       2       2       3         ST-220 complex       1       2       2       3         ST-220 complex       1       2       3       1         ST-220 complex/Cluster A3       1       3       1       1         ST-220 complex/Cluster A3       1       1       1         ST-260 complex       2       1       1       1         ST-260 complex       2       1       1       3         ST-324 complex/Subgroup IV       8       5       1       6         ST-4144 complex/Subgroup IV       8       5       1       6         ST-40 complex       1       2       2       1         ST-40 complex/Subgroup IV       8 <td>ST-162 complex     5     1     2     3       ST-167 complex     1     1     2     3       ST-174 complex     1     1     1     3       ST-176 complex     1     1     1     3       ST-176 complex     1     1     1     1       ST-176 complex     2     2     2       ST-176 complex     3     2     2       ST-26 complex     1     1     1       ST-260 complex     2     1     1       ST-260 complex     1     1     1       ST-360 complex     1     1     1       ST-41 complex/Lineage 3     5     1     6       ST-53 complex/Subgroup III     6     2     2       ST-50 complex     1     1     1       Total     28     27     1     3       St-64 complex/Subgroup III     6     2     2       ST-50 complex     1     1     1       Total     28     27     1     3&lt;</td> <td></td> <td>.5</td> <td>16</td> <td>5</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	ST-162 complex     5     1     2     3       ST-167 complex     1     1     2     3       ST-174 complex     1     1     1     3       ST-176 complex     1     1     1     3       ST-176 complex     1     1     1     1       ST-176 complex     2     2     2       ST-176 complex     3     2     2       ST-26 complex     1     1     1       ST-260 complex     2     1     1       ST-260 complex     1     1     1       ST-360 complex     1     1     1       ST-41 complex/Lineage 3     5     1     6       ST-53 complex/Subgroup III     6     2     2       ST-50 complex     1     1     1       Total     28     27     1     3       St-64 complex/Subgroup III     6     2     2       ST-50 complex     1     1     1       Total     28     27     1     3<		.5	16	5								
ST-167 complex       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I	ST-167 complex       1       2       3         ST-174 complex       1       1       1       3         ST-176 complex       3       1       1       3         ST-180 complex       3       2       2       2         ST-180 complex       2       2       2       2         ST-200 complex       2       2       2       2         ST-220 complex       1       1       1       1         ST-220 complex       2       1       3       1         ST-220 complex       2       1       3       1         ST-220 complex       1       1       1       1         ST-340 complex       1       1       1       1         ST-410 complex/inege 3       1       0       6       6         ST-400 complex       1       1       1       1       1         Todi       28       27       1       3       98       1												
ST-174 complex       1       1       1       1       1       3         ST-178 complex       3       1       1       1       1       1         ST-18 complex       3       2       2       2       3         ST-198 complex       2       2       2       2         ST-22 complex       1       2       2       2         ST-22 complex       1       1       1       1         ST-22 complex/Cluster A3       1       1       1       1         ST-254 complex       1       1       1       1       1         ST-254 complex       1       1       1       1       1       1         ST-254 complex       2       1       1       1       1       1         ST-256 complex       1       1       1       3       3       3         ST-32 complex/ET-5 complex       1       1       1       1       5       5       1       1         ST-4 complex/subgroup IV       8       4       4       1       5       1       5       6         ST-50 complex       1       5       1       6       5       5	ST-174 complex     1     1     1     1     3       ST-178 complex     1     1     1     1       ST-180 complex     3     2     3       ST-198 complex     2     2     2       ST-226 complex     1     1     1       ST-226 complex     2     1     1       ST-226 complex     1     1     1       ST-226 complex     2     1     1       ST-269 complex     2     1     1       ST-260 complex     1     1     1       ST-40 complex/subgroup/ll     6     2     2       ST-61 complex/subgroup/ll     6     2     2       ST-620 complex     1     1     1       ST-61 complex/subgroup/ll     6     2     2       ST-620 complex     1     1     1       Total			-		1			2				
ST-178 complex       1       1       1       1         ST-18 complex       3       2       3       3       3         ST-18 complex       3       2       3       3       3         ST-18 complex       2       2       3       3         ST-98 complex       1       2       2       2         ST-22 complex       1       2       1       1         ST-22 complex       1       1       1         ST-22 complex       1       1       1         ST-22 complex       1       1       1         ST-26 complex       1       1       1       1         ST-26 complex       2       1       1       3         ST-32 complex/ET-5 complex       6       1       1       3         ST-34 complex/Lineage 3       5       1       6       3         ST-41/4 complex/Lineage 3       5       1       6         ST-53 complex       1       2       2       3         ST-60 complex       1       2       2       3         ST-60 complex       1       4       3       1         Total       28       27	ST-178 complex     1     1       ST-18 complex     3     2     3       ST-98 complex     2     2       ST-22 complex     2     2       ST-22 complex     1     2       ST-22 complex     1     1       ST-25 complex     1     1       ST-26 complex     1     1       ST-26 complex     1     1       ST-26 complex     1     1       ST-26 complex     1     1       ST-32 complex/ET-5 complex     6     6       ST-32 complex/ET-5 complex     1     1       ST-35 complex     1     1       ST-46 complex     1     1       ST-46 complex     1     1       ST-46 complex     1     1       ST-46 complex     1     1       ST-47 complex     1     1       ST-40 complex     1     1       Total     28     27     1       ST-40 complex     1				1		1						
ST-18 complex       3       2       3         ST-198 complex       2       2       2         ST-26 complex       1       2       1         ST-256 complex       1       1       1         ST-256 complex       1       1       1         ST-254 complex       1       1       1         ST-254 complex       2       1       1         ST-254 complex       2       1       1         ST-254 complex       2       1       1         ST-256 complex       2       1       1         ST-256 complex       2       1       1         ST-32 complex(ET-5 complex       6       1       1         ST-334 complex       1       1       1         ST-4 complex/subgroup IV       8       8         ST-441 complex/subgroup IV       8       6         ST-401 complex       1       1       6         ST-50 complex       1       2       2         ST-600 complex       1       1       1         Total       28       27       17       3       15       1         Total       28       27       17       <	ST-18 complex     3     2     2       ST-196 complex     2     2       ST-226 complex     1     1       ST-226 complex     1     1       ST-226 complex     1     1       ST-226 complex     2     1       ST-226 complex     2     1       ST-226 complex     2     1       ST-226 complex     2     1       ST-226 complex     1     1       ST-226 complex     2     1       ST-236 complex     1     1       ST-32 complex/Unleage 3     5     1       ST-401 complex/Unleage 3     5     1       ST-50 complex/Subgroup II     6       ST-51 complex/Subgroup II     6       ST-52 complex/Subgroup II     6       ST-53 complex     1     1       Total     28     27     1       Total     28     27     1     1       St-53 complex     1     1     1       St-60 complex     1     1     1       St-60 complex     1     1     1       St-73 complex     1     1     1       St-60 complex     1     1     1       St-60 complex     1     1     1									1			
ST-22 complex       1       2       2         ST-226 complex       1       1       1         ST-254 complex       1       1       1         ST-256 complex       2       1       1         ST-256 complex       2       1       1         ST-256 complex       2       1       3         ST-256 complex       2       1       3         ST-35 complex       1       1       1         ST-35 complex       1       1       1         ST-36 complex       1       1       1         ST-41/4c complex/Lineage 3       5       1       6         ST-41/4c complex/subgroup IV       8       4       1         ST-50 complex       1       5       1         ST-53 complex       1       2       2         ST-60 complex       1       4       1         Total       28       27       17       3       15       3       1       3         ST-60 complex       1       1       1       1       1       1       1	ST-22 complex     1     1       ST-226 complex     1     1       ST-254 complex     1     1       ST-269 complex     2     1       ST-2769 complex     1     1       ST-2769 complex     1     1       ST-334 complex     1     1       ST-40 complex/subgroup IV     8     6       ST-440 complex/subgroup III     6       ST-461 complex     1     1       ST-50 complex     1     1       ST-60 complex     1     1       Total     28     27     17       ST-60 complex     1     1       Total     28     27       St-60 complex     1     1       Total     28     27       St-60 complex     1     1       St-60 complex     1       St-60			3						3			
ST-226 complex       1       1       1         ST-23 complex/Cluster A3       1       1       1         ST-254 complex       1       1       1         ST-269 complex       2       1       1       3         ST-260 complex       2       1       6       3         ST-32 complex/ET-5 complex       6       6       1       5         ST-32 complex/ET-5 complex       6       1       6       1         ST-35 complex       1       6       1       1       1         ST-414 complex/subgroup IV       8       6       1       6       1         ST-414 complex/subgroup IV       8       6       1       6       1         ST-51 complex       1       6       1       6       1       6         ST-51 complex/subgroup III       6       6       1       1       6         ST-53 complex       1       2       2       2       2         ST-60 complex       1       1       1       1       1         Total       28       27       17       3       15       3       1       3         ST-60 complex       1	ST-226 complex       1       1         ST-232 complex/Cluster A3       1       1         ST-269 complex       2       1       3         ST-269 complex       2       1       3         ST-260 complex       2       1       3         ST-260 complex       2       1       3         ST-32 complex/Cluster A3       6       6         ST-34 complex       1       1         ST-34 complex       1       1         ST-44 complex/Lineage 3       5       1       6         ST-54 complex       2       2       2         ST-50 complex       1       1       1         Total       28       27       17       3       15       3       1       98         wnload: Tab-delimited text   Excel format	ST-198 complex				2							
ST-23 complex/Cluster A3       I       I       I       I         ST-254 complex       I       I       I       I         ST-256 complex/ET-5 complex       C       I       I       I         ST-334 complex/ET-5 complex       G       I       I       I         ST-334 complex/LT-5 complex       G       I       I       I         ST-35 complex       I       I       I       I         ST-45 complex/burgoup IV       8       I       I       I         ST-41/44 complex/Lineage 3       5       I       I       I         ST-60 complex       I       I       I       I         ST-53 complex       I       I       I       I         ST-60 complex       I       I       I       I         Total       28       27       I7       3       I5       J       J         Total       28       27       I7       3       I5       J       J       J	ST-23 complex/Cluster A3       1       1       1         ST-254 complex       2       1       3         ST-269 complex       6       3         ST-32 complex/E1-5 complex       6       1         ST-32 complex/E1-5 complex       1       1         ST-334 complex       1       1         ST-35 complex/bigroup IV       8       1         ST-41/44 complex/bigroup IV       8       1         ST-55 complex/bigroup IV       6       1         ST-56 complex       1       1         ST-56 complex/bigroup IV       6       1         ST-50 complex       1       1         Total       28       27       17         Total       28       27       1         Total       28       27       1         Mindad: Tab-delimited text   Excel format       1       1         windad: Tab-delimited text   Excel format       1       1         Values						2			2			
ST-254 complex       1       1       1         ST-269 complex       2       1       3         ST-32 complex/ET-5 complex       6       6       6         ST-334 complex/ET-5 complex       1       1       1         ST-35 complex       1       1       1         ST-35 complex       1       1       1         ST-40 complex/Lineage 3       5       1       6         ST-51 complex       1       6       1         ST-51 complex/Lineage 3       5       1       6         ST-51 complex/Subgroup IV       6       6         ST-51 complex       1       6         ST-53 complex       2       2         ST-60 complex       1       1         Total       28       27       17       3       1       3	ST-254 complex       1       1       1         ST-269 complex       2       1       3         ST-269 complex       2       1       3         ST-32 complexUF15 complex       1       1         ST-34 complex       1       1         ST-35 complex       1       1         ST-35 complex       1       1         ST-4 complexSubgroup IV       8       6         ST-461 complex       1       1         ST-55 complex       1       6         ST-461 complex       1       1         ST-55 complex       2       2         ST-60 complex       1       1         Total       28       27       17       3       15       3       1       98         wnload: Tab-delimited text   Excel format       1       1       1       1       1       1         wnload: Tab-delimited text   Excel format       -       -       -       -       -       -         wnload: Tab-delimited text   Excel format       -       -       -       -       -       -         ''harts       -       -       -       -       -       -       -       -       <			1									
ST-269 complex       2       1       3         ST-32 complex/ET-5 complex       6       6       6         ST-334 complex       1       1       1         ST-35 complex       1       1       1         ST-35 complex       1       1       1         ST-45 complex/subgroup IV       8       8         ST-41/44 complex/Lineage 3       5       1       6         ST-41 complex/subgroup IV       8       6       8         ST-41 complex/subgroup IV       8       6       8         ST-41 complex/subgroup IV       6       1       6         ST-41 complex/subgroup IV       6       1       6         ST-53 complex/subgroup III       6       2       2         ST-60 complex       1       1       1         Total       28       27       17       3       15       3       1       3	ST-269 complex     2     1     3       ST-32 complex     6     6       ST-32 complex     1     1       ST-35 complex     1     1       ST-35 complex     1     8       ST-401 complex/Lineage 3     5     1       ST-401 complex     1     1       ST-50 complex     1     6       ST-50 complex     1     6       ST-50 complex     1     6       ST-50 complex     1     1       Total     28     27     17       Stronplex     1     1								1				
ST-32 complex/ET-5 complex       6       6       6         ST-334 complex       1       6       1         ST-35 complex       1       1       1         ST-45 complex/subgroup IV       8       8       8         ST-41/44 complex/lineage 3       5       1       6         ST-41/44 complex/subgroup IV       8       1       1         ST-41/44 complex/subgroup IV       8       1       1         ST-451 complex       1       1       6         ST-50 complex       1       2       1         ST-50 complex       1       1       1         Total       28       27       17       3       15       3	ST-32 complex       6       6       1         ST-33 complex       1       1       1         ST-35 complex/subgroup IV       8       8         ST-41/44 complex/subgroup IV       8       6         ST-410 complex       1       6         ST-50 complex       1       6         ST-50 complex       1       1         Total       28       27       17       3       15       3       1       98         ownload: Tab-delimited text   Excel format       Percentages         Values       Percentages												
ST-334 complex       1       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I       I	ST-334 complex       1       1         ST-35 complex       1       1         ST-35 complex/subgroup IV       8       8         ST-41/44 complex/Lineage 3       5       1       6         ST-41/44 complex/Lineage 3       5       1       6         ST-41/44 complex/Lineage 3       5       1       6         ST-461 complex       1       1       1         ST-5 complex       1       6         ST-5 complex       2       2         ST-60 complex       1       1         Total       28       27       1         Total       28       27       1       1         str-50 complex       1       1       1         Total       28       27       17       3       15       1       98         swnload: Tab-delimited text   Excel format       Percentages         values       Percentages					1							
ST-35 complex       1       1       1         ST-4 complex/subgroup IV       8       8       8         ST-41/44 complex/Lineage 3       5       1       6         ST-41 complex/subgroup IV       1       6       1         ST-41 complex/subgroup IV       6       1       6         ST-41 complex/subgroup IV       6       1       6         ST-5c complex/subgroup IV       6       2       6         ST-53 complex       2       2       2         ST-60 complex       1       1       1         Total       28       27       17       3       15       3       1       98	ST-35 complex     1     1       ST-4 complex/subgroup IV     8     8       ST-41/44 complex/lineage 3     5     1     6       ST-461 complex/subgroup III     6     6       ST-461 complex     1     1       ST-56 complex     1     6       ST-56 complex     2     2       ST-60 complex     1     1       Total     28     27     17       Total     28     27     17     3       ownload: Tab-delimited text   Excel format												
ST-4 complex/subgroup IV     8     6       ST-41/44 complex/Lineage 3     5     1     6       ST-41/44 complex/Lineage 3     5     1     6       ST-41/64 complex/Lineage 3     5     1     6       ST-41/64 complex/Lineage 3     5     1     6       ST-41/64 complex/Lineage 3     6     1       ST-50 complex     6     2     6       ST-50 complex     7     2     2       Total     28     27     17     3     15     3     1	ST-4 complex/subgroup IV     8     6       ST-41/44 complex/subgroup IV     8     6       ST-41/44 complex/subgroup IV     5     1     6       ST-41 complex/subgroup IV     1     6       ST-52 complex/subgroup IV     6     2       ST-60 complex     1     1       ST-50 complex     2     2       ST-60 complex     1     1       Total     28     27     17       Total     28     27     17       Str-50 complex     1     1       Total     28     27       Str-60 complex     1     1       Total     28     27       Str-60 complex     1     3       Str-60 complex     1     98					1							
ST-414 complex/Lineage 3       5       1       6         ST-461 complex       1       1       1         ST-50 complex       6       1         ST-50 complex       6       2       6         ST-50 complex       6       2       6         ST-50 complex       1       2       2       2         ST-60 complex       1       1       1         Total       28       27       17       3       15       3       1       98	ST-41/44 complex     1     6       ST-461 complex     1     1       ST-55 complex/subgroup III     6       ST-53 complex     2     2       ST-60 complex     1     1       Total     28     27     17       Stromplex     1     1     1		8										
ST-461 complex       1       1       1         ST-55 complex/subgroup III       6       6         ST-53 complex       2       2         ST-60 complex       1       1         Total       28       27       17       3       15       3       1       98	ST-461 complex       1       1       6         ST-5 complex       2       2         ST-50 complex       1       1         Total       28       27         Strick to enlarge.			5		1							
ST-5 complex/subgroup III     6     6       ST-53 complex     2     2       ST-60 complex     1     1       Total     28     27     17     3     15     3     1     3     1     98	ST-5 complex/subgroup III     6     2     2       ST-50 complex     2     2       ST-60 complex     1     1       Total     28     27     17     3     15     3     1     3     1     98												
ST-53 complex     2     2     2       ST-60 complex     1     1     1       Total     28     27     17     3     15     3     1     3     1     98	ST-53 complex     1     1     1       Total     28     27     17     3     15     3     1     98		6										
Total 28 27 17 3 15 3 1 3 1 98	Total     28     27     17     3     15     3     1     98       wnload: Tab-delimited text   Excel format					2				2			
	wwnload: Tab-delimited text   Excel format	ST-60 complex			1					1			
wnload: Tab-delimited text   Excel format	harts lick to enlarge. — Values		28	27 17	/ 3	15	3	1	3 1	1 98			
wnload: Tab-delimited text   Excel format	harts lick to enlarge. — Values — Percentages												
	harts lick to enlarge. — Values — Percentages	wnload: Tab-delimited text   E	cel fo	rmat									
	lick to enlarge. 												
	lick to enlarge. Values — Percentages — Per	harte											
harte	Values Percentages	nants											
harts		ick to enlarge.											
		Values			Perce	ntanee							
ick to enlarge.				2 52	rence	mayes							
lick to enlarge. - Values		Da			8								
ick to enlarge. - Values			·		2								
ick to enlarge. - Values		- 1 🖬 📕 📕 👘			2								
lick to enlarge. 				( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )	- <b>-</b>				1.1	1			
lick to enlarge. 							1.1.1.1.1.1		Mar				

The table values can be exported in a format suitable for copying in to a spreadsheet by clicking 'Download as tabdelimited text' underneath the table.

**Note:** The job will be submitted to the offline job queue if the query returns 10,000 or more isolates. In this case, the buttons to reverse the axes or to change whether values or percentages are shown will not be available.

# 12.4 Scheme and allele breakdown

The scheme and allele breakdown plugin displays the frequency of each allele and scheme field (e.g. ST or clonal complex).

The function can be selected for the whole database by clicking the 'Scheme and allele breakdown' link on the main contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence de database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a popul database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a popul of database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a popul of database.  Option settings Submissions Search database - peruse all records. Browse database. Browse database. Browse database - peruse all records. Browse database - peruse all records. Browse database - peruse all records. Browse database. Browse database - peruse all records. Browse database. Browse database - peruse all records. Browse database -	ation sample. ssions nation
<ul> <li>Breakdown</li> <li>Single field</li> <li>Two field</li> <li>Unique combinations</li> <li>Scheme and alleles</li> <li>Sequence bin</li> </ul> Export dataset <ul> <li>Contigs</li> <li>Sequences - XMFA / concatenated FASTA formats</li> <li>Codon usage</li> <li>Presence/absence status of loci</li> <li>Genome comparator</li> <li>BLAST</li> </ul> Miscellaneous <ul> <li>Description of database fields</li> </ul>	

Alternatively, a breakdown can be displayed of the dataset returned from a query by clicking the 'Schemes/alleles' button in the Breakdown list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

1	A4/M1027	B1; Z1001	USA	1937	invasive (unspecified/other)	Neisseria men	ingitidis	A	4	ST-4 complex/subgroup IV	5-2	10	F1-5
2	120M	B35; Z1035	Pakistan		meningitis and septicaemia			A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
3	M00242905		UK	2000	invasive (unspecified/other)	Neisseria men	ingitidis	В	1099		19	15	
4	M1027	B43; Z1043	USA	1937	invasive (unspecified/other)	Neisseria men	ingitidis	Α	4	ST-4 complex/subgroup IV			
5	M00240227		UK	2000	invasive (unspecified/other)	Neisseria men	ingitidis	В	1100	ST-32 complex/ET-5 complex	7	16	
6	M00282207		UK	2000	invasive (unspecified/other)	Neisseria men	ingitidis	W	1101	ST-22 complex			
7	7891	B54; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria men	ingitidis	Α	5	ST-5 complex/subgroup III	20	9	F3-1
8	M00242007		UK	2000	invasive (unspecified/other)	Neisseria men	ingitidis	В	1102	ST-18 complex		14	
9	0021/84		Czech Republic	1984	invasive (unspecified/other)	Neisseria men	ingitidis	W	114	ST-22 complex			
10	6748	B73; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria men	ingitidis	Α	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
11	129	B92; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria men	ingitidis	A	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
12	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria men	ingitidis	в	1015	ST-32 complex/ET-5 complex	7	16	
13	139M	B99; Z1099	Philippines	1968		Neisseria men	ingitidis	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
14	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria men	ingitidis	Х	117			14	
15	1		Germany	1999	carrier	Neisseria men	ingitidis	E	864				
16	2		Germany	1999	carrier	Neisseria men	ingitidis	в	854	ST-18 complex			
17	3		Germany	1999	carrier	Neisseria men	ingitidis	W	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria men	ingitidis	в	19	ST-18 complex			
19		B213; Z1213	Ghana		invasive (unspecified/other)	Neisseria men	ingitidis	Α	4	ST-4 complex/subgroup IV	7	13-1	F1-5
20	5		Germany	1999		Neisseria men		NG	198	ST-198 complex			
21	6		Germany	1999		Neisseria men		NG	198	ST-198 complex			
22	7		Germany	1999		Neisseria men		E	60	ST-60 complex			
23	8		Germany	1999		Neisseria men	-	В		ST-32 complex/ET-5 complex			
24		B227; Z1227	Denmark		invasive (unspecified/other)			A	5	ST-5 complex/subgroup III	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria men	ingitidis	В	930	ST-334 complex			
25 In	9 alysis tools akdown: Fi	i: elds Two Fi	Germany	1999 Polyr		Neisseria men Neisseria men Schemes/allele	ingitidis		930	SI-5 complex/Subgroup III ST-334 complex	5-1	g	F3-1

A scheme tree is shown. Select any combination of schemes to analyse. Click 'Select'.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
	Help 🗹
Scheme field and allele breakdown of dataset	Show options
Select schemes or groups of schemes within the tree. A breakdown of the individual fields and loci belonging to these schemes will then be performed.  Select schemes  Action  Action  MLST  Finetyping antigens  Antigen genes  Action  Actio	

A table showing the number of unique values for each locus and scheme field will be displayed.

PuliMLST		n: Isolate f	fields   Scl	e/ST   List neme/allele ons   Profile			ions						
													Help 🖉
Scheme fie	eld and	allele	e brea	kdowi	n of d	latase	t						Show options
							_						
Select schemes of	r groups of s	chemes w	/ithin the tr	ee. A break	down of ti	he individu	al fields	and loci	belongin	g to these :	chemes w	ill then be pe	erformed.
						Action —			-	-			
All loci													
					â Ľ	Select							
	netic Informa	tion Proce	ssina		=								
	nomic island		Joonig										
	eage Schem												
	tabolism				-								
					÷								
MLST													
ſ	Fields			All	eles								
Field name U	nique values	Analyse	Locus U	nique allele	s Analys	e Downloa	ad						
ST	547		abcZ	64	<b>B</b>	) 🛃							
clonal complex	43		adk	56		) 👱							
			aroE	80	¢								
			fumC	83									
			gdh	77									
			pdhC	82									
			pgm	68	¢								

A detailed display of allele or field frequencies can be displayed by clicking the appropriate 'Breakdown' button.

PubMLST	Breakdown:	ch   Browse   Prof Isolate fields   So nts   Home   Opti	heme/alleles			ns				
										Help 🖓
Scheme fie	eld and a	allele brea	akdown	of da	taset					Show options
e- 🛄 💭 Ger	psule netic Informatio nomic islands eage Schemes tabolism	n Processing		Ac	individual tion — lect	fields ar	nd loci beloi	nging to these s	chemes will then	be performed.
MLST										
F	Fields		Alle	les						
Field name U	nique values A	Analyse Locus (	<b>Jnique allele</b> s	Analyse	Download	1				
ST	547	🕒 abcZ	64		<b>±</b>					
clonal complex	43	🖨 adk	56		<b>±</b>					
		aroE	80	¢	1					
		fumC	83		_ <u></u>					
		gdh	77		*					
		pdhC	82 68	¢	*					
		pgm	00							

The sorting of the table can be changed by clicking the appropriate header - this toggles between ascending and descending order.

TIDIMLST / Breakdown:	Isolate fiel	e   Profile/ST   List ds   Scheme/allele e   Options   Profile	s   Publications s/sequences definitions	
				Help 🗗
cheme field and a	allele l	breakdow	n of dataset	Show optic
requency table				
atal: 005 is slates				
otal: 985 isolates				
clonal_complex 🔶 Fr	equency 🖨	Percentage 🔶		
ST-11 complex/ET-37 complex	186	18.88		
ST-41/44 complex/Lineage 3	99	10.05		
ST-1 complex/subgroup I/II	63	6.40		
ST-8 complex/Cluster A4	60	6.09		
ST-22 complex	53	5.38		
ST-32 complex/ET-5 complex	51	5.18		
ST-4 complex/subgroup IV	37	3.76		
ST-269 complex	31	3.15		
ST-5 complex/subgroup III	31	3.15		
ST-23 complex/Cluster A3	25	2.54		
ST-60 complex	24	2.44		
ST-167 complex	18	1.83		
ST-18 complex	14	1.42		
ST-213 complex	13	1.32		
ST-35 complex	12	1.22		
ST-162 complex	9	0.91		
ST-174 complex	9	0.91		
ST-53 complex	8	0.81		
ST-92 complex	8	0.81		
ST-106 complex	7	0.71		
ST-254 complex	7	0.71		
ST-103 complex	6	0.61		
ST-334 complex	6	0.61		
ST-198 complex	5	0.51		
ST-865 complex	5	0.51		
ST-178 complex	4	0.41		
ST-364 complex	4	0.41		
ST-37 complex	4	0.41		
ST-1157 complex	3	0.30		

The table values can be exported in a format suitable for copying in to a spreadsheet by clicking the 'Tab-delimited text' button.

ST-92 complex ST-106 complex ST-254 complex ST-103 complex ST-103 complex ST-334 complex ST-344 complex ST-198 complex ST-198 complex ST-364 complex ST-37 complex ST-37 complex ST-1157 complex ST-116 complex ST-1494 complex (lactamica) ST-4240/6688 complex ST-624 complex (lactamica) ST-1540 complex (lactamica) ST-1540 complex (lactamica) ST-613 complex ST-613 complex ST-613 complex ST-226 complex ST-264 complex (lactamica) ST-226 complex ST-264 complex (lactamica) ST-226 complex ST-264 complex (lactamica) ST-264 complex (lactamica) ST-595 complex (lactamica) ST-595 complex (lactamica) ST-640 complex (lactamica) ST-640 complex (lactamica)	8 7 7 6 5 5 4 4 4 3 3 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2	0.81 0.71 0.71 0.61 0.61 0.51 0.51 0.41 0.41 0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20
ST-106 complexST-254 complexST-103 complexST-103 complexST-34 complexST-198 complexST-865 complexST-865 complexST-78 complexST-364 complexST-37 complexST-1157 complexST-116 complexST-1494 complex (lactamica)ST-624 complex (lactamica)ST-1136 complexST-1540 complex (lactamica)ST-212 complexST-613 complex (lactamica)ST-750 complexST-226 complexST-595 complex (lactamica)ST-595 complex (lactamica)ST-640 complex (lactamica)	7 6 5 5 4 4 4 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 1 1	0.71 0.61 0.61 0.51 0.51 0.41 0.41 0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20
ST-254 complexST-103 complexST-334 complexST-34 complexST-198 complexST-865 complexST-865 complexST-178 complexST-37 complexST-37 complexST-1157 complexST-116 complexST-1494 complex (lactamica)ST-624 complex (lactamica)ST-1136 complexST-1540 complex (lactamica)ST-212 complexST-613 complex (lactamica)ST-505 complexST-295 complex (lactamica)ST-595 complex (lactamica)ST-640 complex (lactamica)	7 6 5 5 4 4 4 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2	0.71 0.61 0.51 0.51 0.41 0.41 0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20
ST-103 complex ST-334 complex ST-198 complex ST-198 complex ST-865 complex ST-178 complex ST-364 complex ST-37 complex ST-1157 complex ST-116 complex ST-1494 complex (lactamica) ST-4240/6688 complex ST-624 complex (lactamica) ST-136 complex ST-641 complex ST-613 complex (lactamica) ST-750 complex ST-613 complex ST-226 complex ST-282 complex ST-595 complex (lactamica) ST-595 complex (lactamica) ST-640 complex (lactamica)	6 6 5 4 4 4 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2	0.61 0.51 0.51 0.41 0.41 0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20
ST-334 complexST-198 complexST-198 complexST-198 complexST-1865 complexST-364 complexST-37 complexST-37 complexST-1157 complexST-116 complexST-116 complexST-1494 complex (lactamica)ST-624 complex (lactamica)ST-1136 complexST-1540 complex (lactamica)ST-1540 complex (lactamica)ST-212 complexST-613 complex (lactamica)ST-505 complexST-282 complexST-595 complex (lactamica)ST-640 complex (lactamica)	6 5 4 4 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2	0.61 0.51 0.51 0.41 0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20
ST-198 complex ST-865 complex ST-364 complex ST-37 complex ST-37 complex ST-1157 complex ST-116 complex ST-1494 complex (lactamica) ST-4240/6688 complex ST-624 complex (lactamica) ST-1136 complex ST-614 complex (lactamica) ST-212 complex ST-613 complex (lactamica) ST-750 complex ST-613 complex (lactamica) ST-226 complex ST-282 complex ST-595 complex (lactamica) ST-595 complex (lactamica)	5 5 4 4 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2	0.51 0.51 0.41 0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20
ST-865 complex ST-178 complex ST-364 complex ST-37 complex ST-1157 complex ST-116 complex ST-116 complex ST-1494 complex (lactamica) ST-4240/6688 complex ST-624 complex (lactamica) ST-1136 complex ST-614 complex (lactamica) ST-212 complex ST-613 complex (lactamica) ST-750 complex ST-613 complex (lactamica) ST-226 complex ST-282 complex ST-595 complex (lactamica) ST-595 complex (lactamica) ST-640 complex (lactamica)	5 4 4 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2	0.51 0.41 0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20
ST-178 complex ST-364 complex ST-37 complex ST-1157 complex ST-1167 complex ST-1166 complex T-1494 complex (lactamica) ST-4240/6688 complex T-624 complex (lactamica) ST-4240/6688 complex ST-4240/6688 complex ST-4240/6688 complex ST-4240/6688 complex ST-1136 complex (lactamica) ST-1136 complex ST-212 complex ST-212 complex ST-261 complex ST-260 complex ST-282 complex ST-282 complex ST-295 complex (lactamica) ST-640 complex (lactamica)	4 4 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2	0.41 0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20
ST-364 complex ST-37 complex ST-1157 complex ST-116 complex ST-1494 complex (lactamica) ST-4240/6688 complex T-624 complex (lactamica) ST-1136 complex F-1540 complex (lactamica) ST-212 complex ST-461 complex T-613 complex (lactamica) ST-750 complex ST-226 complex ST-282 complex ST-282 complex ST-282 complex ST-285 complex (lactamica) T-640 complex (lactamica)	4 3 3 3 3 2 2 2 2 2 2 2 2 2 2 1 1	0.41 0.30 0.30 0.30 0.30 0.30 0.30 0.20 0.20
ST-37 complex ST-1157 complex ST-1157 complex ST-116 complex T-1494 complex (lactamica) ST-4240/6688 complex ST-624 complex (lactamica) ST-1136 complex T-1540 complex (lactamica) ST-212 complex ST-461 complex ST-613 complex (lactamica) ST-750 complex ST-266 complex ST-282 complex ST-282 complex ST-295 complex (lactamica) ST-640 complex (lactamica)	4 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 1 1	0.41 0.30 0.30 0.30 0.30 0.30 0.20 0.20 0.20
ST-1157 complex ST-116 complex ST-1494 complex (lactamica) ST-4240/6688 complex ST-624 complex (lactamica) ST-1136 complex ST-1136 complex ST-212 complex ST-212 complex ST-613 complex (lactamica) ST-750 complex ST-26 complex ST-282 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	3 3 3 2 2 2 2 2 2 2 2 2 2 2 1 1	0.30 0.30 0.30 0.30 0.20 0.20 0.20 0.20
ST-116 complex I-1494 complex (lactamica) ST-4240/6688 complex T-624 complex (lactamica) ST-1136 complex I-1540 complex (lactamica) ST-212 complex ST-461 complex T-613 complex (lactamica) ST-750 complex ST-226 complex ST-282 complex ST-282 complex I-595 complex (lactamica) T-640 complex (lactamica)	3 3 2 2 2 2 2 2 2 2 2 2 2 1 1	0.30 0.30 0.30 0.20 0.20 0.20 0.20 0.20
T-1494 complex (lactamica) ST-4240/6688 complex T-624 complex (lactamica) ST-1136 complex T-1540 complex (lactamica) ST-212 complex ST-461 complex T-613 complex (lactamica) ST-750 complex ST-226 complex ST-282 complex ST-282 complex ST-295 complex (lactamica) T-640 complex (lactamica)	3 2 2 2 2 2 2 2 2 2 1 1	0.30 0.30 0.20 0.20 0.20 0.20 0.20 0.20
ST-4240/6688 complex ST-624 complex (lactamica) ST-1136 complex T-1540 complex (lactamica) ST-212 complex ST-613 complex (lactamica) ST-750 complex ST-26 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	3 2 2 2 2 2 2 2 2 2 1 1	0.30 0.20 0.20 0.20 0.20 0.20 0.20 0.20
ST-624 complex (lactamica) ST-1136 complex ST-1540 complex (lactamica) ST-212 complex ST-461 complex ST-613 complex (lactamica) ST-750 complex ST-226 complex ST-282 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	3 2 2 2 2 2 2 2 2 1 1	0.20 0.20 0.20 0.20 0.20 0.20
ST-1136 complex ST-1540 complex (lactamica) ST-212 complex ST-461 complex ST-613 complex (lactamica) ST-750 complex ST-226 complex ST-282 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	2 2 2 2 2 2 2 1 1	0.20 0.20 0.20 0.20 0.20 0.20
ST-1540 complex (lactamica) ST-212 complex ST-461 complex ST-613 complex (lactamica) ST-750 complex ST-226 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	2 2 2 2 2 1 1	0.20 0.20 0.20 0.20 0.20 0.20
ST-212 complex ST-461 complex ST-613 complex (lactamica) ST-750 complex ST-226 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	2 2 2 2 1 1	0.20 0.20 0.20 0.20
ST-461 complex ST-613 complex (lactamica) ST-750 complex ST-226 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	2 2 1 1	0.20
ST-750 complex ST-226 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	2 1 1	0.20
ST-226 complex ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	1 1	
ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)	1	0.10
ST-282 complex ST-595 complex (lactamica) ST-640 complex (lactamica)		
ST-595 complex (lactamica) ST-640 complex (lactamica)		0.10
ST-640 complex (lactamica)	1	0.10
	1	0.10
harts	1	
lick to enlarge.		
Pie chart		Bar chart

You can also download the sequences for alleles designated in the dataset for the loci belonging to the scheme by clicking the appropriate 'Download' button in the first results table.

PubMLST	Breakdow	arch   Browse   Profil /n: Isolate fields   Sch ntents   Home   Optio	eme/allel	es   Publications es/sequences definitions	
					Help 🖓
Scheme fie	ld and	allele brea	kdow	n of dataset	Show options
Select schemes or performed.	groups of s	chemes within the tr	ee. A brea	kdown of the individual fields and loci belongi	ng to these schemes will then be
	s			Action	
🛓 📄 🎧 All loci				▲ Select	
b- Car		ation Processing			
	nomic island	-			
	eage Schem	nes			
Met	abolism -			<b>T</b>	
MLST					
	ields			leles	
Field name U				es Analyse Download	
ST	547	🚯 abcZ	64		
clonal complex	43	🚯 adk	56		
		aroE	80		
		fumC	83 77		
		gdh pdhC	82		
		pane	68		

#### Sequences will be served in FASTA format in order of frequency.

>2
TTTGATACCGTTGCCGAAGGTTTGGGTGAAATTCGCGATTTATTGCGCCGTTACCACCGC
GTCGGCCATGAGTTGGAAAACGGTTCGGGTGAGGCTTTGTTGAAAGAACTCAACGAATTA
CAACTTGAAATCGAAGCGAAGGACGGCTGGAAGCTGGATGCGGCAGTCAAGCAGACTTTG
GGGGAACTCGGTTTGCCGGAAAAACGAAAAAATCGGCAACCTTTCCGGCGGTCAGAAAAAG
CGTGTCGCCTTGGCGCAGGCTTGGGTGCAGAAGCCCGACGTATTGCTGCTGGACGAACCG
ACCAACCATTTGGATATCGACGCGATTATTTGGCTGGAAAATCTGCTCAAAGCGTTTGAA
GGCAGCTTGGTTGTGATTACCCACGACCGCCGTTTTTTGGACAATATCGCCACGCGGATT
GTCGAACTCGATC
>1
TTTGATACTGTTGCCGAAGGTTTGGGCGAAATTCGCGATTTATTGCGCCGTTATCATCAT
GTCAGCCATGAGTTGGAAAATGGTTCGAGTGAGGCCTTATTGAAAGAGCTCAACGAATTG
CAACTTGAGATCGAAGCGAAGGACGGCTGGAAGTTGGATGCGGCGGTGAAGCAGACTTTG
GGCGAACTCGGTTTGCCGGAAAAACGAAAAAATCGGCAACCTCTCCGGCGGTCAGAAAAAG
CGCGTCGCCTTGGCGCAGGCTTGGGTGCAGAAGCCCGACGTATTGCTGCTCGATGAACCG
ACCAACCATTTGGACATCGACGCGATTATTTGGTTGGAAAACCTGCTCAAAGCGTTTGAA
GGCAGCCTGGTTGTGATTACCCACGACCGCCGTTTTTTGGACAATATCGCCACGCGGATT
GTCGAACTCGATC
>4
TTTGATACCGTTGCCGAAGGTTTGGGCGAAATTCGTGATTTATTGCGCCGTTATCATCAT
GTCAGCCATGAGTTGGAAAATGGTTCGAGTGAGGCTTTGTTGAAAGAACTCAACGAATTG
CAACTTGAAATCGAAGCGAAGGACGGCTGGAAACTGGATGCGGCAGTCAAGCAGACTTTG
GGGGAACTCGGTTTGCCGGAAAATGAAAAAATCGGCAACCTTTCCGGCGGTCAGAAAAAG
CGCGTCGCCTTGGCTCAGGCTTGGGTGCAAAAGCCCGACGTATTGCTGCTGGACGAGCCG
ACCAACCATTTGGATATCGACGCGATTATTTGGCTGGAAAATCTGCTCAAAGCGTTTGAA

```
GGCAGCTTGGTTGTGATTACCCACGACCGCCGTTTTTTGGACAATATCGCCACGCGGATT
GTCGAACTCGATC
```

# 12.5 Sequence bin breakdown

The sequence bin breakdown plugin calculates statistics based on the number and length of contigs in the sequence bin as well as the number of loci tagged for an isolate record.

The function can be selected by clicking the 'Sequence bin' link on the Breakdown section of the main contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences defi	nitions
there is at least one corresponding isolate deposited here. Any isolate may be s Query database • Search database - advanced queries. • Browse database - peruse all records.	represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database submitted to this database and consequently it should be noted that it does not represent a population sample. Option settings • Set general options - including isolate table field handling. • Set display and query options for locus, schemes or scheme fields.
<ul> <li>Search by combinations of loci (profiles) - including partial matching.</li> <li>List query - find isolates by matching a field to an entered list.</li> <li>Projects - main projects defined in database.</li> </ul>	General information • Isolates: 34221 • Last updated: 2015-07-02 • Update history • About BIGSdb
Breakdown     Single field     Two field     Unique combinations     Scheme and alleles     Publications     Sequence bin	Analysis - Codon usage - Presence/absence status of loci - RASTA formats - RASTA formats - BLAST - Miscellaneous - Description of database fields - Description of database fields

Alternatively, it can be accessed following a query by clicking the 'Sequence bin' button in the Breakdown list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

Select the isolate records to analyse - these will be pre-selected if you accessed the plugin following a query. You

can also select loci and/or schemes which will be used to calculate the totals and percentages of loci designated and tagged. This may be useful as a guide to assembly quality if you use a scheme of core loci where a good assembly would be expected to include all member loci. To determine the total of all loci designated or tagged, click 'All loci' in the scheme tree.

There is also an option to determine the mean G+C content of the sequence bin of each isolate.

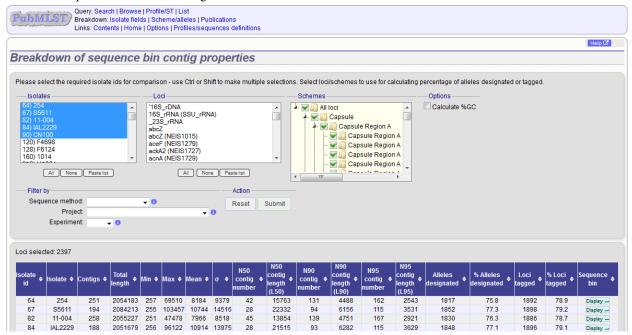
	elds   Scheme/alleles   Publication elds   Scheme/alleles   Publication ne   Options   Profiles/sequences				Help 🖉
eakdown of sequenc	e bin contig prope	erties			
ease select the required isolate ids for c	comparison - use Ctrl or Shift to m	ake multiple selectio	ns. Select loci/schemes to use f	or calculating percentage of	f alleles designated or tagged.
-Isolates	— — Loci —			Options	
64) 254 67) 55611 84) 14.2229 90) CN100 120) F4698 128) F6124 160) 1014 All None Pastelist	116S_rDNA 16S_rRNA (SSU_rRNA) _23S_rRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) ackA2 (NEIS1727) acnA (NEIS1729) All None Pa	Tiste list	All Ioci     Capsule     Genetic Informatio     Genetic Informatio     Genomic islands     Lineage Schemes     Metabolism     Pilin     Tilin	=	e %GC
Filter by		Action	-		
Sequence method: Project:	- 0 - 0	Reset Submit	J		
Experiment: 🚽 🜖					

If there are fewer than 100 isolates selected, the table will be generated immediately. Otherwise it will be submitted to the job queue.

A table of sequence bin stats will be generated.

~ . .

.



You can choose to export the data in tab-delimited text or Excel formats by clicking the appropriate link at the bottom of the table.

67	S5611 19	4 2084213	255	103457	10744	14516	28	22332	94	6156	115	3531	1852	77.3	1898	79.2	Display
82	11-004 25	8 2055227	251	47478	7966	8518	45	13854	139	4751	167	2921	1830	76.3	1886	78.7	Display
84	IAL2229 18	8 2051679	256	96122	10914	13975	28	21515	93	6282	115	3629	1848	77.1	1896	79.1	Display -
90	CN100 23	6 2118544	261	73069	8977	11101	36	19065	119	4570	149	2839	1848	77.1	1898	79.2	Display ->
Dor Click on the Number of	Opwnload in tab-delimited text format     Opwnload in Excel format Click on the following charts to enlarge Number of contigs Total length Mean contig length Contig lengths																
1	an: 225.4; o: 32.5			n: 207276		T	Overall me		σ. 1396.7		hilling	196.0; σ: 114 Ταιτικάτατα 15					

Sequence bin records can also be accessed by clicking the 'Display' button for each row of the table.

_	Experimer	nt ,	0			-												
ci selected	d: 2397																	
olate id 🕈 Isi	solate 🗢	Contigs 🗢	Total length 🕈	Min ¢	Max ¢	Mean ¢		N50 contig <b>4</b> number		N90 contig ¢ number	N90 contig length (L90)	N95 contig number	N95 contig length (L95)	Alleles designated 🕈	% Alleles designated 🕈	Loci tagged 🕈	% Loci tagged \$	Sequence bin
64	254	251	2054183	257	69510	8184	9379	42	15763	131	4488	162	2543	1817	75.8	1892	78.9	Display
67	S5611	194	2084213	255	103457	10744	14516	28	22332	94	6156	115	3531	1852	77.3	1898	79.2	Display
82	11-004	258	2055227	251	47478	7966	8518	45	13854	139	4751	167	2921	1830	76.3	1886	78.7	Display
84 I/	IAL2229	188	2051679	256	96122	10914	13975	28	21515	93	6282	115	3629	1848	77.1	1896	79.1	Display
90 0	CN100	236	2118544	261	73069	8977	11101	36	19065	119	4570	149	2839	1848	77.1	1898	79.2	Display

# 12.6 Genome comparator

Genome Comparator is an optional plugin that can be enabled for specific databases. It is used to compare whole genome data of isolates within the database using either the database defined loci or the coding sequences of an annotated genome as the comparator.

Output is equivalent to a whole genome MLST profile, a distance matrix calculated based on allelic differences and a NeighborNet graph generated from this distance matrix.

Genome Comparator can be accessed on databases where it is enabled from the contents page by clicking the 'Genome Comparator' link.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Schemeralleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in t database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it doe	
<ul> <li>Query database</li> <li>Search database - advanced queries.</li> <li>Browse database - peruse all records.</li> <li>Search by combinations of loci (profiles) - including partial matching.</li> <li>List query - find isolates by matching a field to an entered list.</li> <li>Projects - main projects defined in database.</li> </ul>	Submissions • Manage submissions General information • Isolates: 34221 • Last updated: 2015-07-02 • Update history • About BIGSdb
Breakdown       Export       Analysis       Miscellane         • Single field       • Export dataset       • Contigs       • Codon usage       • Description         • Two field       • Ontigue combinations       • Sequences - XMFA / concatenated FASTA formats       • Codon usage       • Description         • Publications       • Sequence bin       • Sequence Security       • Description       • Description	ous of database fields

Alternatively, it can be accessed following a query by clicking the 'Genome Comparator' button at the bottom of the results table. Isolates with sequence data returned in the query will be automatically selected within the Genome Comparator interface.

L									
19025 M10 240481	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	1831		22	9	F3-3
19026 M10 240482	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	3754	ST-41/44 complex/Lineage 3	7-2	4	F5-1
19027 M10 240484	UK	2010	invasive (unspecified/other) Neisseria meningitidis	В	1097	ST-41/44 complex/Lineage 3	17-1	23	F1-5
19028 M10 240485	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	275	ST-269 complex	22	9	F5-12
19029 M10 240487	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	1049	ST-269 complex	19-1	15-11	F5-1
19030 M10 240489	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	461	ST-461 complex	19-2	13-1	F3-9
19031 M10 240490	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	1161	ST-269 complex	22	9	F5-12
19032 M10 240498	UK	2010	invasive (unspecified/other) Neisseria meningitidis	В	41	ST-41/44 complex/Lineage 3	7-2	4	F1-5
19958 M10 240476	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	9812	ST-213 complex	22	14	F5-5
19959 M10 240499	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	275	ST-269 complex	22	9	F5-12
19960 M10 240500	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	41	ST-41/44 complex/Lineage 3	7-2	4	F1-5
19961 M10 240502	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	340	ST-41/44 complex/Lineage 3	7-2	4	F1-5
19962 M10 240503	UK	2010	invasive (unspecified/other) Neisseria meningitidis	Y	23	ST-23 complex/Cluster A3	5-1	2-2	F1-96
19963 M10 240505	UK	2010	invasive (unspecified/other) Neisseria meningitidis	Y	1655	ST-23 complex/Cluster A3	5-1	10-10	F4-1
19964 M10 240507	UK	2010	invasive (unspecified/other) Neisseria meningitidis	Y	183	ST-23 complex/Cluster A3	21	16-5	deleted
19965 M10 240508	UK	2010	invasive (unspecified/other) Neisseria meningitidis	В	1575		7-2	13-1	F1-7
19966 M10 240511	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	4713		22	9	F5-12
19967 M10 240512	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	41	ST-41/44 complex/Lineage 3	7-1	1	F1-5
19968 M10 240514	UK	2010	invasive (unspecified/other) Neisseria meningitidis	W	11	ST-11 complex/ET-37 complex	5	2	F1-1
19969 M10 240515	UK	2010	invasive (unspecified/other) Neisseria meningitidis	в	269	ST-269 complex	5-1	2-2	F5-1
19970 M10 240520	UK	2010	invasive (unspecified/other) Neisseria meningitidis	Y	1655	ST-23 complex/Cluster A3	5-1	10-1	F4-1
Analysis: Presence/Absen	10e Ge	Codons anome ( Bequent	Comparator BLAST	Publ	ications	Sequence bin			

## 12.6.1 Analysis using defined loci

Select the isolate genomes that you wish to analyse and then either the loci from the list or a set of schemes. Press submit.

	Profile/ST   List s   Scheme/alleles   Publications   Options   Profiles/sequences definitions   Data	base submissions		
				Help 🖉 Toggle: i
Genome Comparator - Ne	eisseria PubMLST			Show options
	ti for comparison - use ctrl or shift to make multi , you can enter the accession number for an anr			defined in schemes by selecting the
Isolates 655) 890326 659) A22 660) 71/94 661) 860800 662) 2837 6633 2839 664) 2838 655) 2845 All None Pastelist	Loci '16S_rDNA 16S_rRNA (SSU_rRNA) 233_rRNA abc2 abc2 (NEIS1015) aceF (NEIS1279) achd2 (NEIS1727) achd2 (NEIS1727) All None Paste list	Include in identifiers Sci icountry country region year age yr age mth sex disease source	Ineage Schemes	
Reference genome Enter accession number: I or choose annotated genome: or upload Genbank/EMBL/FASTA file: Browse No file selected.	Parameters / options Min % identity: 70 • 7 Min % alignment 50 • 7 BLASTN word size: 15 • 7 Use TBLASTX [] V Use tagged designations if available 7 Disable HTML output 7	Distance matrix calculation With incomplete loci: © Completely exclude from analysis Treat as distinct allele © Ignore in pairwise comparison Exclude paralogous loci © paralogous in all isolates © paralogous in any isolate	Alignments Produce alignments Align at loci (not only variable) Align at loci (not only variable) Aligner: MAFFT •	Core genome analysis Core threshold (%): 90 ▼ [ Calculate mean distances []
Filter by Sequence method: Project Experiment:	Action Reset Sub	mit		

The job will be submitted to the job queue and will start running shortly. Click the link to follow the job progress and view the output.

Publications Links: Contents Home   Options   Profiles/st List Links: Contents Home   Options   Profiles/sequences definitions   Database submissions	
	Toggle: i
Genome Comparator - Neisseria PubMLST	Show options
This analysis has been submitted to the job queue.	
Please be aware that this job may take a long time depending on the number of comparisons and how busy the server is.	
Follow the progress of this job and view the output.	

There will be a series of tables displaying variable loci, colour-coded to indicate allelic differences. Finally, there will be links to a distance matrix which can be loaded in to SplitsTree for further analysis and to a NeighborNet chart showing relatedness of isolates. Due to processing constraints on the web server, this NeighborNet is only calculated if 200 or fewer genomes are selected for analysis, but this can be generated in the stand-alone version of SplitsTree using the distance matrix if required.

	Browse   Profile/ST   List					
	ate fields   Scheme/alleles   F   Home   Options   Profiles/se		se submissions			
ob status viewer						
Status						
Job id: BIGSdb_27748_14054	410063_94241					
Submit time: 2014-07-15 08:41:03 Status: finished						
Start time: 2014-07-15 08:41:05 Progress: 100%						
Stop time: 2014-07-15 08:41:22 Total time: 17 seconds						
Dutput						
Analysis against defined loci						
llele numbers are used where these	e have been defined, otherwis	e sequences will be marked	as 'New#1, 'New#2' etc. I	lissing alleles are mark	ked as 'X'. Truncated alleles (located at end	of contig) are marked as 'T
Locus 644 (L93/4286) 662 (2837) ( abcZ 2 2	663 (2839) 664 (2838) 665 ( 2 2 2	2845) 666 (2843) 667 (2842 2 2 2	) 669 (2846) 670 (2840) 7 2	671 (2844) 672 (2847) 2 2	698 (FAM18)	
adk 3 3 aroE 4 19			3 3 4 19	3 3 19 19	3 4	
fumC 3 3 gdh 8 8	3 3 2 8 8 8		3 3 8 8	3 3 8 8	3 8	
pdhC 4 4 pgm 6 6	4 4 4 6 6 6	6 6			4 6	
oci with sequence differences	among isolates:					
ariable loci: 4						
Locus 644 (L93/4286) 662 (2837) 6 abcZ 2 2	663 (2839) 664 (2838) 665 (	2845) 666 (2843) 667 (2842	) 669 (2846) 670 (2840) 7 2	671 (2844) 672 (2847)	698 (FAM18)	
aroE 4 19 fumC 3 3	4 4 4 3 3 2		4 19 3 3	19 19 3 3	4	
pdhC 4 4	4 4 4	6 6	4 4	4 4	4	
latches: 3 Locus 644 (L93/4286) 662 (2837) 6 adk 3 3 gdh 8 8 pgm 6 6	663 (2839) 664 (2838) 665 ( 3 3 3 3 8 8 8 6 6 6 6	3 3 8 8 8	) 669 (2846) 670 (2840) 3 3 8 8 6 6	671 (2844)         672 (2847)           3         3           8         8           6         6	698 (FAM18) 3 8 6	
Inique strains						
Jnique strains: 5						
Strain 1         Strain 2         Strain 3           544 (L93/4286) 662 (2837) 666 (284)         563 (2839)         670 (2840) 667 (284)           564 (2838)         671 (2844)         564 (2838)         571 (2844)           598 (FAM18)         672 (2847)         572 (2847)         572 (2847)	3) 665 (2845) 669 (2846)					
Text output file     Excel format     Distance matrix (Nexus format     Splits graph (Neighbour-net, F		plitsTree. Distances between	taxa are calculated as th	e number of loci with dif	fferent allele sequences	
	(click to enlarge)	nd in Inkeenne er etherur der	graphics aditors			
Splits graph (Neighbour-net, S     Locus presence frequency     Locus presence frequency ch		ed in missage of other vector	graphics editors			
	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,					
Tar file containing output files	(click to enlarge)					
ease note that job results will remain						

#### 12.6.2 Analysis using annotated reference genome

Select the isolate genomes that you wish to analyse and then either enter a Genbank accession number for the reference genome, or select from the list of reference genomes (this list will only be present if the administrator has *set it up*). Selecting reference genomes will hide the locus and scheme selection forms.

	Profile/ST   List s   Scheme/alleles   Publications   Options   Profiles/sequences definitions   Data	ubase submissions		
				Help 🖉 Toggle: 1
Genome Comparator - Ne	eisseria PubMLST			Show options
	ci for comparison - use ctri or shift to make multi you can enter the accession number for an anr Include in identifiers Include in identifiers Country region year epidemiological year age yr age yr age mth sex disease source			defined in schemes by selecting the
Reference genome Enter accession number: or choose annotated genome: FAM18 (Nm) v or uproad Genoan/EMBL/FASTA file: Browse No file selected. [] Filter by Sequence method: v [] Project Experiment. v []	Parameters / options       Min % identity: 70 • [       Min % alignment: 50 • [       BLASTN word size: 15 • [       Use TBLASTN [       Use TBLASTN [       Osable HTML output [       Action       Reset       Subr	Distance matrix calculation With incomplete loci: © Completely exclude from analysis Treat as distinct allele Ignore in pairwise comparison Ø Exclude paralogous loci © paralogous in all isolates paralogous in any isolate mit	Alignments Produce alignments [ Include ref sequences in alignment Align all loc (not only variable) Aligner: MAFFT	Core genome analysis Core threshold (%): 90  ▼ [] Calculate mean distances []

Output is similar to when comparing against defined loci, but this time every coding sequence in the annotated reference will be BLASTed against the selected genomes. Because allele designations are not defined, the allele found in the reference genome is designated allele 1, the next different sequence is allele 2 etc.

b stat	us viewer															
tatus																
Job id: ubmit time: Status: Start time: Progress: Stop time:	2014-07-15 09:12:58															
Total time:	24 minutes and 24 seconds															
1 - C	reference genome															
accession	_ *															
version	1															
type	dna															
length	2194961															
description	Neisseria meningitidis serogroup C FAM18 complete genome.															
descriptior	Neisseria meningitidis serogroup C FAM18 complete genome.															
descriptior oding regio	Neisseria meningitidis serogroup C FAM18 complete genome.															
descriptior ding regio	Nelsseria meningitidis serogroup C FAM18 complete genome. 18 1975															
lescription ding regio loci	Neisseria meningitidis serogroup C FAM18 complete genome.	X'. Truncated	alleles (loca	ated at end o	f contig) are	marked	as T.									
lescriptior ding regio loci ch unique a	Neisseria meningilidis serogroup C FAM18 complete genome. 1975 Allele is defined a number starting at 1. Missing alleles are marked as "				f contig) are 644			664	665	666	667	669	670	671	672	69
lescriptior ding regio loci	Nelsseria meningitidis serogroup C FAM18 complete genome. 18 1975	X'. Truncated Sequence length		ated at end o Reference genome		662	663	664 (2838)	665 (2845)	666 (2843)	667 (2842)	669 (2846)	670 (2840)	671 (2844)	672 (2847)	
lescription ding regio loci ch unique a Locus (C   envA	Neisseria meningilidis serogroup C FAM18 complete genome. 1975 Allele is defined a number starting at 1. Missing alleles are marked as "	Sequence	Genome	Reference	644	662	663									
lescriptior ding regio loci ch unique a Locus (C   envA   IMC0001	Neisseria meningitidis serogroup C FAM18 complete genome. 1975 Illele is defined a number starting at 1. Missing alleles are marked as " Product UDP-3-O-[3-hydroxymyristoyi] N-acetylglucosmine deacetylase	Sequence length 924	Genome position 1261	Reference genome 1	644 (L93/4286) 1	662 (2837) 1	663	(2838) 1	(2845) 1	(2843) 1	(2842) 1	(2846) 1	(2840) 1	(2844) 1	(2847) 1	
lescriptior ding regio loci ch unique a Locus (C   envA   IMC0001 piIS1	Nelsseria meningitidis serogroup C FAM18 complete genome. ns 1975 Allele is defined a number starting at 1. Missing alleles are marked as 7 Product	Sequence length	Genome position	Reference genome	644	662 (2837)	663		(2845)		(2842)	(2846)		(2844)	(2847)	
escriptior ding regio loci th unique a Locus (C   envA   MC0001 pilS1   MC0002	Neisseria meningitidis serogroup C FAM18 complete genome. Is 1975 Illele is defined a number starting at 1. Missing alleles are marked as " Product UDP-3-O-[3-hydroxymyristoy] N-acetylglucosmine deacetylase pilin (fragment)	Sequence length 924 291	Genome position 1261 3341	Reference genome 1 1	644 (L93/4286) 1 1	662 (2837) 1 1	663 (2839) 1 1	(2838) 1 1	(2845) 1 1	(2843) 1 1	(2842) 1 1	(2846) 1 1	(2840) 1 1	(2844) 1 1	(2847) 1 1	(FAN
lescription ding regio loci th unique a Locus (C   envA   IMC0001 pilS1   IMC0002 pilS2	Neisseria meningitidis serogroup C FAM18 complete genome. 1975 Illele is defined a number starting at 1. Missing alleles are marked as " Product UDP-3-O-[3-hydroxymyristoyi] N-acetylglucosmine deacetylase	Sequence length 924	Genome position 1261	Reference genome 1	644 (L93/4286) 1	662 (2837) 1	663	(2838) 1	(2845) 1	(2843) 1	(2842) 1	(2846) 1	(2840) 1	(2844) 1	(2847) 1	(FAN
lescription ding regio loci ch unique a Locus xC   envA   IMC0001 piIS1   IMC0002 piIS2   IMC0003	Neisseria meningitidis serogroup C FAM18 complete genome. Is 1975 Illele is defined a number starting at 1. Missing alleles are marked as " Product UDP-3-O-[3-hydroxymyristoyi] N-acetylglucosmine deacetylase pilin (fragment) truncaled pilin	Sequence length 924 291 366	Genome position           1261           3341           3675	Reference genome 1 1 1	644 (L93/4286) 1 1 2	662 (2837) 1 1 2	663 (2839) 1 1 2	(2838) 1 1 2	(2845) 1 1 2	(2843) 1 1 2	(2842) 1 1 2	(2846) 1 1 2	(2840) 1 1 2	(2844) 1 1 2	(2847) 1 1 2	(FAN
lescription ding regio loci ch unique a Locus kC   envA   IMC0001 pilS1   IMC0002 pilS2   pilS2   IMC0003 fbp	Neisseria meningitidis serogroup C FAM18 complete genome. Is 1975 Illele is defined a number starting at 1. Missing alleles are marked as " Product UDP-3-O-[3-hydroxymyristoy] N-acetylglucosmine deacetylase pilin (fragment)	Sequence length 924 291	Genome position 1261 3341	Reference genome 1 1	644 (L93/4286) 1 1	662 (2837) 1 1	663 (2839) 1 1	(2838) 1 1	(2845) 1 1	(2843) 1 1	(2842) 1 1	(2846) 1 1	(2840) 1 1	(2844) 1 1	(2847) 1 1	(FAN
Iescription ding regio loci ch unique a Locus xC   envA   JMC0001 pilS1   JMC0002 pilS2   JMC0003 fbp   JMC0004	Neisseria meningilidis serogroup C FAM18 complete genome. Illele is defined a number starting at 1. Missing alleles are marked as " Product UDP-3-0-[3-hydroxymyristoyi] N-acetylglucosmine deacetylase pilin (fragment) truncated pilin peplidyl-prolyl cis-trans isomerase	Sequence length           924           291           366           330	Genome position 1261 3341 3675 4069	Reference genome 1 1 1 1 1	644 (L93/4286) 1 1 2 2	662 (2837) 1 1 2 2	663 (2839) 1 1 2 2	(2838) 1 1 2 2	(2845) 1 2 2	(2843) 1 2 2	(2842) 1 2 2	(2846) 1 2 T	(2840) 1 2 2 2	(2844) 1 2 2	(2847) 1 2 2	(FAN 1 1 1
Iescription ding regio loci ch unique a kC   enA   JMC0001 pilS1   JMC0002 pilS2   JMC0003 fbp   JMC0005	Neisseria meningitidis serogroup C FAM18 complete genome. Is 1975 Illele is defined a number starting at 1. Missing alleles are marked as " Product UDP-3-O-[3-hydroxymyristoy] N-acelylglucosmine deacelylase pillin (fragment) truncated pillin peptidyl-prolyl cis-trans isomerase putative membrane protein	Sequence length 924 291 366 330 219	Genome position           1261           3341           3675           4069           4476	Reference genome 1 1 1 1 1 1	644 (L93/4286) 1 2 2 2	662 (2837) 1 2 2 3	663 (2839) 1 2 2 3	(2838) 1 2 2 3	(2845) 1 2 2 3	(2843) 1 2 2 3	(2842) 1 2 2 3	(2846) 1 2 2 T 4	(2840) 1 2 2 3	(2844) 1 2 2 5	(2847) 1 2 2 3	(FAN 1 1 1
description dding regio l loci ch unique a locus xC   envA   vMC0001 pilS1   vMC0002 pilS2   vMC0003 fbp   vMC0004 vMC0004 vMC0006	Neisseria meningitidis serogroup C FAM18 complete genome. 1975 Illele is defined a number starting at 1. Missing alleles are marked as 7 Product UDP-3-O-[3-hydroxymyristoy] N-acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-protyl cis-trans isomerase putative ghycerate dehydrogenase	Sequence length 924 291 366 330 219 954	Genome position           1261           3341           3675           4069           4476           4816	Reference genome 1 1 1 1 1 1 1 1	644 (L93/4286) 1 2 2 2 2 2	662 (2837) 1 2 2 3 2	663 (2839) 1 1 2 2 3 2	(2838) 1 2 2 3 2 2	(2845) 1 2 2 3 2	(2843) 1 2 2 3 2 2	(2842) 1 2 2 3 2	(2846) 1 2 2 T 4 2	(2840) 1 2 2 3 2	(2844) 1 2 2 5 2	(2847) 1 2 2 3 3 2	69 (FAM 1 1 1 1 1
description oding regio I loci I loci Locus DXC   envA   NMC0001 pilS1   NMC0002 pilS2   NMC0003	Neisseria meningitidis serogroup C FAM18 complete genome. Is 1975 Illele is defined a number starting at 1. Missing alleles are marked as " Product UDP-3-O-[3-hydroxymyristoy] N-acelylglucosmine deacelylase pillin (fragment) truncated pillin peptidyl-prolyl cis-trans isomerase putative membrane protein	Sequence length 924 291 366 330 219	Genome position           1261           3341           3675           4069           4476	Reference genome 1 1 1 1 1 1	644 (L93/4286) 1 2 2 2	662 (2837) 1 2 2 3	663 (2839) 1 2 2 3	(2838) 1 2 2 3	(2845) 1 2 2 3	(2843) 1 2 2 3	(2842) 1 2 2 3	(2846) 1 2 2 T 4	(2840) 1 2 2 3	(2844) 1 2 2 5	(2847) 1 2 2 3	(FAN 1 1 1

### 12.6.3 Include in identifiers fieldset

This selection box allows you to choose which isolate provenance fields will be included in the results table. This does not affect the output of the alignments as taxa names are limited in length by the alignment programs.

-Include in identifiers	
isolate	
country	
region	=
year	
epidemiological year	
age yr	
age mth	
sex	
disease	
source	-

Multiple values can be selected by clicking while holding down Ctrl.

#### 12.6.4 Reference genome fieldset

This section allows you to choose a reference genome to use as the source of comparator sequences.

Reference genome Enter accession number: İ or choose annotated genome: FAM18 (Nm) or upload Genbank/EMBL/FASTA file: No file selected. Browse... i

There are three possibilities here:

- 1. Enter accession number Enter a Genbank accession number of an annotated reference and Genome Comparator will automatically retrieve this from Genbank.
- 2. Select from list The administrator may have selected some genomes to offer for comparison. If these are present, simply select from the list.
- 3. Upload genome Click 'Browse' and upload your own reference. This can either be in Genbank, EMBL or FASTA format. Ensure that the filename ends in the appropriate file extension (.gb, .embl, .fas) so that it is recognized.

### 12.6.5 Parameters/options fieldset

This section allows you to modify BLAST parameters. This affects sensitivity and speed.

-Parameters / options	
Min % identity:	70 👻 👔
Min % alignment:	50 👻 👔
BLASTN word size:	15 👻 👔
Use TBLASTX i	
🗹 Use tagged desigr	nations if available 👔
🗖 Disable HTML out	out 👔

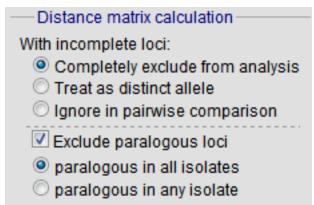
- Min % identity This sets the threshold identity that a matching sequence has to be in order to be considered (default: 70%). Only the best match is used.
- Min % alignment This sets the percentage of the length of reference allele sequence that the alignment has to cover in order to be considered (default: 50%).
- BLASTN word size This is the length of the initial identical match that BLAST requires before extending a match (default: 20). Increasing this value improves speed at the expense of sensitivity. The default value gives good results in most cases. The default setting used to be 15 but the new default of 20 is almost as good (there was 1 difference among 2000 loci in a test run) but the analysis runs twice as fast.
- Use TBLASTX This compares the six-frame translation of your nucleotide query sequence against the sixframe translation of the contig sequences. Sequences will be classed as identical if they result in the same translated sequence even if the nucleotide sequence is different. This is significantly slower than using BLASTN.

Additionally, two other options are available in this fieldset:

- Use tagged designations When analysing using defined loci, Genome Comparator can use the designations stored within the database (this is the default). This is much quicker since it doesn't need to run BLAST against these sequences. If a designation is missing, BLAST will be run for that locus anyway.
- Disable HTML output If running Genome Comparator against a large number of genomes, the resulting table may get so large that your web browser struggles to render it properly and may use up too much memory on your computer. Clicking this button prevents this output this output is not required for further analysis since everything present in it is also generated in Excel format at the end. HTML output is automatically disabled when more than 150 genomes are analysed.

### 12.6.6 Distance matrix calculation fieldset

This section provides options for the treatment of incomplete and paralogous loci when generating the distance matrix.



For incomplete loci, i.e. those that continue beyond the end of a contig so are incomplete you can:

- Completely exclude from analysis Any locus that is incomplete in at least one isolate will be removed from the analysis completely. Using this option means that if there is one bad genome with a lot of incomplete sequences in your analysis, a large proportion of the loci may not be used to calculate distances.
- Treat as a distinct allele This treats all incomplete sequences as a specific allele 'I'. This varies from any other allele, but all incomplete sequences will be treated as though they were identical.
- Ignore in pairwise comparison (default) This is probably the best option. In this case, incomplete alleles are only excluded from the analysis when comparing the particular isolate that has it. Other isolates with different alleles will be properly included. The effect of this option will be to shorten the distances of isolates with poorly sequenced genomes with the others.

Paralogous loci, i.e. those with multiple good matches, can be excluded from the analysis (default). This is the safest option since there is no guarantee that differences seen between isolates at paralogous loci are real if the alternative matches are equally good. NB: Loci are also only classed as paralogous when the alternative matches identify different sequences, otherwise multiple contigs of the same sequence region would result in false positives.

When paralogous loci are excluded, there are two further options:

- Exclude when paralogous in all isolates (default). Loci are only classed as paralogous when there are multiple hits in every genome (except if a genome is missing the locus entirely, in which case that genome is ignored in the calculation). This is generally the option that you will want to use with the default BLAST parameters since you can often expect multiple hits even when loci are not paralogous if you have used relaxed thresholds.
- Exclude when paralogous in any isolate. Unless you use stringent BLAST thresholds, this is likely to overestimate the number of paralogous loci, but may be useful if you are specifically looking for them.

#### 12.6.7 Alignments fieldset

This section enables you to choose to produce alignments of the sequences identified.

Available options are:

- Produce alignments Selecting this will produce the alignment files, as well as XMFA and FASTA outputs of aligned sequences. This will result in the analysis taking approximately twice as long to run.
- Include ref sequences in alignment When doing analysis using an annotated reference, selecting this will include the reference sequence in the alignment files.
- Align all loci By default, only loci that vary among the isolates are aligned. You may however wish to align all if you would like the resultant XMFA and FASTA files to include all coding sequences.
- Aligner There are currently two choices of alignment algorithm (provided they have both been installed)
  - MAFFT (default) This is the preferred option as it is significantly quicker than MUSCLE, uses less memory, and produces comparable results.

MUSCLE - This was originally the only choice. It is still included to enable previous analyses to be re-run
and compared but it is recommended that MAFFT is used otherwise.

#### 12.6.8 Core genome analysis fieldset

This section enables you to modify the inclusion threshold used to calculate whether or not a locus is part of the core genome (of the dataset).

— Core genome analysis —	
Core threshold (%): 90 👻	i
Calculate mean distances	i

The default setting of 90% means that a locus is counted as core if it appears within 90% or more of the genomes in the dataset.

There is also an option to calculate the mean distance among sequences of the loci. Selecting this will also select the option to produce alignments.

### 12.6.9 Filter fieldset

This section allows you to further filter your collection of isolates and the contigs to include.

-Filter by		
Sequence method:	▼ i	
Project:		<b>√</b> [i]
Experiment:	▼ [i]	

Available options are:

- Sequence method Choose to only analyse contigs that have been generated using a particular method. This depends on the method being set when the contigs were uploaded.
- Project Only include isolates belonging to the chosen project. This enables you to select all isolates and filter to a project.
- Experiment Contig files can belong to an experiment. How this is used can vary between databases, but this enables you to only include contigs from a particular experiment.

### 12.6.10 Understanding the output

#### **Distance matrix**

The distance matrix is simply a count of the number of loci that differ between each pair of isolates. It is generated in NEXUS format which can be used as the input file for SplitsTree. This can be used to generate NeighborNet, Split decomposition graphs and trees offline. If 200 isolates or fewer are included in the analysis, a Neighbor network is automatically generated from this distance matrix.

#### **Unique strains**

The table of unique strains is a list of isolates that are identical at every locus. Every isolate is likely to be classed as unique if a whole genome analysis is performed, but with a constrained set of loci, such as those for MLST, this will group isolates that are indistinguishable at that level of resolution.

# 12.7 **BLAST**

The BLAST plugin enables you to BLAST a sequence against any of the genomes in the database, displaying a table of matches and extracting matching sequences.

The function can be accessed by selecting the 'BLAST' link on the Analysis section of the main contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences defin	litions
	represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition may be submitted to this database and consequently it should be noted that it does not represent a population sample.
Query database  Search database - advanced queries.  Browse database - peruse all records.  Search by combinations of loci (profiles) - including partial matching.  List query - find isolates by matching a field to an entered list.  Projects - main projects defined in database.	<ul> <li>Option settings         <ul> <li>Set general options - including isolate table field handling.</li> <li>Set display and query options for locus, schemes or scheme fields.</li> </ul> </li> <li>General information         <ul> <li>Isolates: 34221</li> <li>Last updated: 2015-07-02</li> <li>Update history</li> <li>About BIGSdb</li> </ul> </li> </ul>
Breakdown     Single field     Two field     Unique combinations     Scheme and alleles     Publications     Sequence bin     Sequence bin	Analysis Codon usage Presence/absence status of loci BLAST ASTA formats Analysis Miscellaneous Description of database fields

Alternatively, it can be accessed following a query by clicking the 'BLAST' button in the Analysis list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

240227 282207 891 B54; Z1054 242007 21/84 748 B73; Z1073 129 B92; Z1092	UK Czech Republic Canada	2000 1975 2000 1984	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitio Neisseria meningitio Neisseria meningitio	tis W tis A	1100 1101 5	ST-32 complex/ET-5 complex ST-22 complex ST-5 complex/subgroup III	7	16 9	
891 B54; Z1054 242007 21/84 748 B73; Z1073 129 B92; Z1092	Finland UK Czech Republic Canada	1975 2000 1984	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitio Neisseria meningitio	lis A			20	q	
242007 21/84 748 B73; Z1073 129 B92; Z1092	UK Czech Republic Canada	2000 1984	invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitio		5	ST 5 complex/cubaroup III	20	9	
21/84 748 B73; Z1073 129 B92; Z1092	Czech Republic Canada	1984	invasive (unspecified/other)		tic D		31-3 complex subgroup in			F3-1
748 B73; Z1073 129 B92; Z1092	Canada				is D	1102	ST-18 complex		14	
129 B92; Z1092		1971		Neisseria meningitio	lis W	114	ST-22 complex			
	Cormony		invasive (unspecified/other)	Neisseria meningitio	lis A	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
		1964	invasive (unspecified/other)	Neisseria meningitio	tis A	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
90/89	Czech Republic	1989	invasive (unspecified/other)	Neisseria meningitio	tis B	1015	ST-32 complex/ET-5 complex	7	16	
39M B99; Z1099	Philippines	1968		Neisseria meningitio	tis A	1	ST-1 complex/subgroup I/I	5-2	10	F5-1
20/95	Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitio	tis X	117			14	
1	Germany	1999	carrier	Neisseria meningitio	lis E	864				
2	Germany	1999	carrier	Neisseria meningitio	tis B	854	ST-18 complex			
3	Germany	1999	carrier	Neisseria meningitio	lis W	174	ST-174 complex			
4	Germany	1999	carrier	Neisseria meningitio	lis B	19	ST-18 complex			
3131 B213; Z121	3 Ghana	1973	invasive (unspecified/other)	Neisseria meningitio	tis A	4	ST-4 complex/subgroup IV	7	13-1	F1-5
5	Germany	1999	carrier	Neisseria meningitio	lis NG	198	ST-198 complex			
6	Germany	1999	carrier	Neisseria meningitio	tis NG	198	ST-198 complex			
7	Germany	1999	carrier	Neisseria meningitio	lis E	60	ST-60 complex			
8	Germany	1999	carrier	Neisseria meningitio	tis B	32	ST-32 complex/ET-5 complex			
4355 B227; Z122	7 Denmark	1974	invasive (unspecified/other)	Neisseria meningitio	lis A	5	ST-5 complex/subgroup III	5-1	9	F3-1
9	Germany	1999	carrier	Neisseria meningitio	tis B	930	ST-334 complex			
3	20095 1 2 3 4 4 5 5 6 7 8 8 355 B227; Z122'	20/95         Czech Republic           1         Germany           2         Germany           3         Germany           4         Germany           5         Germany           6         Germany           7         Germany           8         Germany           9         Germany	20/95         Czech Republic 1995           1         Germany 1999           2         Germany 1999           3         Germany 1999           4         Germany 1999           5         Germany 1999           6         Germany 1999           7         Germany 1999           8         Germany 1999           355         B227, Z1227           9         Germany 1999	Czech Republic 1995         invasive (unspecified/other)           1         Germany         1999         Carrier           2         Germany         1999         carrier           3         Germany         1999         carrier           4         Germany         1999         carrier           131         B213, Z1213         Ghana         1973         invasive (unspecified/other)           5         Germany         1999         carrier           131         B213, Z1213         Ghana         1973         invasive (unspecified/other)           5         Germany         1999         carrier           7         Germany         1999         carrier           8         Germany         1999         carrier           8         Germany         1999         carrier           355         B227, Z1227         Denmark         1974         invasive (unspecified/other)	Czsch Republic         1995         invasive (unspecified/other)         Neisseria meningitic           1         Germany         1999         carrier         Neisseria meningitic           2         Germany         1999         carrier         Neisseria meningitic           3         Germany         1999         carrier         Neisseria meningitic           4         Germany         1999         carrier         Neisseria meningitic           11         B213; Z1213         Ghana         1973         invasive (unspecified/other)         Neisseria meningitic           5         Germany         1999         carrier         Neisseria meningitic           6         Germany         1999         carrier         Neisseria meningitic           7         Germany         1999         carrier         Neisseria meningitic           8         Germany         1999         carrier         Neisseria meningitic           85         B227; Z1227         Denmark         1974         invasive (unspecified/other)         Neisseria meningitic	Czech Republic 1995         invasive (unspecified/other)         Neisseria meningitidis         X           1         Germany         1999         carrier         Neisseria meningitidis         E           2         Germany         1999         carrier         Neisseria meningitidis         B           3         Germany         1999         carrier         Neisseria meningitidis         W           4         Germany         1999         carrier         Neisseria meningitidis         B           11         B213, Z1213         Ghana         1973         invasive (unspecified/other)         Neisseria meningitidis         A           5         Germany         1999         carrier         Neisseria meningitidis         NG           6         Germany         1999         carrier         Neisseria meningitidis         NG           6         Germany         1999         carrier         Neisseria meningitidis         NG           7         Germany         1999         carrier         Neisseria meningitidis         E           8         Germany         1999         carrier         Neisseria meningitidis         B           355         B227, Z1227         Denmark         1974         invasive (unspecified/o	Burger         Czech Republic         1995         invasive (unspecified/other)         Neisseria meningitidis         X         117           1         Germany         1999         carrier         Neisseria meningitidis         E         864           2         Germany         1999         carrier         Neisseria meningitidis         B         854           3         Germany         1999         carrier         Neisseria meningitidis         B         854           4         Germany         1999         carrier         Neisseria meningitidis         B         199           131         B213; Z1213         Ghana         1973 invasive (unspecified/other)         Neisseria meningitidis         A         4           5         Germany         1999         carrier         Neisseria meningitidis         NG         198           6         Germany         1999         carrier         Neisseria meningitidis         NG         198           7         Germany         1999         carrier         Neisseria meningitidis         NG         198           7         Germany         1999         carrier         Neisseria meningitidis         E         60           8         Germany         1999	Czech Republic 1995         invasive (unspecified/other)         Neisseria meningitidis         X         117           1         Germany         1999         carrier         Neisseria meningitidis         E         864           2         Germany         1999         carrier         Neisseria meningitidis         E         864           3         Germany         1999         carrier         Neisseria meningitidis         B         854           4         Germany         1999         carrier         Neisseria meningitidis         B         19         ST-18 complex           131         B213, Z1213         Ghana         1973         invasive (unspecified/other)         Neisseria meningitidis         A         4         ST-18 complex           5         Germany         1999         carrier         Neisseria meningitidis         A         4         ST-18 complex           6         Germany         1999         carrier         Neisseria meningitidis         NG         198         ST-198 complex           7         Germany         1999         carrier         Neisseria meningitidis         E         60         ST-90 complex           8         Germany         1999         carrier         Neisseria meningitidis	20/95     Czech Republic 1995 invasive (unspecified/other) Neisseria meningitidis     X     117       1     Germany     1999     carrier     Neisseria meningitidis     E     864       2     Germany     1999     carrier     Neisseria meningitidis     B     854     ST-18 complex       3     Germany     1999     carrier     Neisseria meningitidis     B     854     ST-178 complex       4     Germany     1999     carrier     Neisseria meningitidis     B     19     ST-174 complex       131     B213, Z1213     Ghana     1973 invasive (unspecified/other)     Neisseria meningitidis     A     4     ST-42 complex/subgroup IV     7       5     Germany     1999     carrier     Neisseria meningitidis     NG     198     ST-198 complex       7     Germany     1999     carrier     Neisseria meningitidis     E     60     ST-90 complex       7     Germany     1999     carrier     Neisseria meningitidis     B     32     ST-90 complex       8     Germany     1999     carrier     Neisseria meningitidis     B     32     ST-90 complex       8     Germany     1999     carrier     Neisseria meningitidis     B     32     ST-32 complex/ET-5 complex	Czech Republic 1995 invasive (unspecified/other)       Neisseria meningitidis       X       117       1       14         1       Germany       1999       carrier       Neisseria meningitidis       E       864       14         2       Germany       1999       carrier       Neisseria meningitidis       E       864       57-18 complex       14         3       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       14         1131       B213, Z1213       Ghana       1973 invasive (unspecified/other)       Neisseria meningitidis       A       4       ST-18 complex       14         1131       B213, Z1213       Ghana       1973 invasive (unspecified/other)       Neisseria meningitidis       A       4       ST-18 complex       14         1131       B213, Z1213       Ghana       1973 invasive (unspecified/other)       Neisseria meningitidis       NG       198       ST-198 complex       14         1131       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       14         114       Germany       1999       carrier       Neisseria meningitidis       E       60       ST-90 complex <td< td=""></td<>

Select the isolate records to analyse - these will be pre-selected if you accessed the plugin following a query. Paste in a sequence to query - this be either a DNA or peptide sequence.

Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definit	tions   Database submissions				
					Toggle: i
BLAST					
67) S5611 SCAASCCTSBATAAAATCSCCAAC S2) 11-004 S4) JAL2229 90) CN100 120) F4698 120) F4698 1203 F6124 SCCAACGCCTSCAASCCAATSCAASCCATS SCCAACCCCAATSCCAASCCATSCAASCCATS SCCAACCCCTTSCAATCCAATCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATSCAASCCATS	GCCGTTGCCGCTTCGGGCGCATTGG GACAICCSCTGGCTGGCAAGCGGC AACGAGCGGGGTGCGTCCATCAIG ACCAIGGTGCGCGCGGCGGTGGTCG GSCAITTCGACTGAACGCGTCGAC CSCCTGTTGGGCGACGCGTCGACG GTACCGGAAAAAATCGACTATTTC	WAAACGCTGSCG ICGCSCTISCGGT ICGCSCTISCGGT ICGCSCTISCG ICGCCCGTIAIC ICGCCCGTAIC ICGCCCGAA	tide sequences can be quer 		
All None	CBIAAMAICUUCIACUAAAAC		source epidemiology	*	
Parameters	Options	Restrict included sequences by		Actio	in
BLASTN word size: 11 👻 👔	Show 0% matches in table	Sequence method:	▼ i	Res	et Submit
BLASTN scoring: reward:2; penalty:-3; gap open:5; gap extend:2 👻 🧵		Project:	-	<b>▼</b> []	
Hits per isolate: 1 -		Experiment 🚽 👔	]		
Flanking length (bp): 100 👻 👔					

#### Click submit.

#### A table of BLAST results will be displayed.

													Toggle
AS	Т												
ase se	elect the requ	uired isolate	ids to BLAST ag	ainst (use ctr	l or shif	ft to make n	nultiple se	lections) and paste	in your qu	ery sequei	nce. Nucleotide	or peptide sequences can be queried.	
Isolat	tes		P	aste sequenc	:e								
) A4/W ) 120M ) 7891 0) 674 1) 129 3) 139 9) S3 4) S43	1 48 9 9 131 355	None	GC: TT AA AC GC	AAGCCTGAAT GGGCGAAAT CCCGACCCAA CATCGGCAT CTACAACCT CTGCGCCGT	TAAAA CAAAA ATGCG GGCGG CTTGC CGGCA	TCGCCAAC TCCCCGAA AAGCGATG GCGCGTCG AATCCATC TCGAACCC	GACATCO AACGAGO ACCATGO GGCAATT CGCCTGI GTACCGO	CCGCTTCGGGCGC GCTGGCTGGCAAG CCGGGTTCGTCCAT TGTGCTGCCAAGT TTCGAGCTGAACGT TGGGCGACCGGTG RAAAAATCGACTA TCGGCTACGAAAA	CGGCCCG CATGCCG GTTCGGC CTATATG CAACAGC TTTCCTG	CGCTGCG GGCAAAG AACGACG CCCGTTA TTCAACG	GGT GTC GTT ATC GAA	country region region region region region region regional set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of the set of	
Parar	meters						Optio	ns		Restrict in	cluded sequence	ces by	
		ze: 11 👻 🚺	1					w 0% matches in t		Sequence			Reset Submit
			⊐ l; penalty:-3; gap	open:5: gap	extend	2 - 1					Project:		i submit
			, pondity, o, gap	oponio, gap	on on o					Ex	operiment:	▼ i	
lanki	lits per isolal ng length (bj e TBLASTX [	p): 100 ·	· i										
lanki Us	ng length (bj e TBLASTX [	p): 100	• I	Mismatches	s Gaps	Seqbin id	Start	End	Orientatio	E-value	Bit score		
Flanki Us late io 1	ng length (bj e TBLASTX [ l lsolate A4/M1027	p): 100	lignment length 465	7	0	180177	19444	19908 extract →	Orientation →	0.0	807		
Tanki Us late id 1 2	ng length (b) e TBLASTX [ l lsolate A4/M1027 120M	p): 100	lignment length 465 465	7 7	0	180177 180583	19444 4782	19908 extract -		0.0 0.0	807 807		
Tankii Us late id 1 2 7	ng length (bj e TBLASTX [ 1 Isolate A4/M1027 120M 7891	p): 100 <b>% identity A</b> 98.49 98.49 98.49 98.49	lignment length 465 465 465	7 7 7	0 0 0	180177 180583 180965	19444 4782 19869	19908 extract	$\stackrel{\rightarrow}{\leftarrow} \stackrel{\leftarrow}{\rightarrow}$	0.0 0.0 0.0	807 807 807		
lanki Us late id 1 2 7 10	ng length (b) e TBLASTX [ i isolate A4/M1027 120M 7891 6748	p): 100 <b>% identity A</b> 98.49 98.49 98.49 98.49 98.49	lignment length 465 465 465 465 465	7 7 7 7	0 0 0 0 0	180177 180583 180965 181186	19444 4782 19869 19181	19908 extract	$\uparrow \downarrow \uparrow \uparrow$	0.0 0.0 0.0 0.0	807 807 807 807		
lanki Us late id 1 2 7 10 11	ng length (b) e TBLASTX [ i isolate A4/M1027 120M 7891 6748 129	<ul> <li>p): 100</li> <li>% identity A 98.49</li> <li>98.49</li> <li>98.49</li> <li>98.49</li> <li>98.49</li> <li>98.49</li> <li>98.49</li> </ul>	lignment length 465 465 465 465 465	7 7 7 7 7 7	0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867	19444 4782 19869 19181 35889	19908 extract	1 4 1 1 1 1	0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807		
Tankii Us ate id 1 2 7 10 11 13	ng length (by e TBLASTX [ A4/M1027 120M 7891 6748 129 139M	<ul> <li>b): 100</li> <li>b): 100</li> <li>b): 100</li> <li>b): 100</li> <li>c): 100</li> <lic): 100<="" li=""> <lic): 100<="" li=""> <lic): 100<="" li=""> <li< td=""><td>lignment length 465 465 465 465 465 465 465</td><td>7 7 7 7 7 7 7</td><td>0 0 0 0 0</td><td>180177 180583 180965 181186 181867 182004</td><td>19444 4782 19869 19181 35889 36775</td><td>19908 extract - 5246 extract - 20333 extract - 19645 extract - 36353 extract - 37239 extract -</td><td>1 1 1 1 1</td><td>0.0 0.0 0.0 0.0 0.0 0.0</td><td>807 807 807 807 807 807 807</td><td></td><td></td></li<></lic):></lic):></lic):></ul>	lignment length 465 465 465 465 465 465 465	7 7 7 7 7 7 7	0 0 0 0 0	180177 180583 180965 181186 181867 182004	19444 4782 19869 19181 35889 36775	19908 extract - 5246 extract - 20333 extract - 19645 extract - 36353 extract - 37239 extract -	1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807		
lankii Us late id 1 2 7 10 11 13 19	ng length (b) e TBLASTX [ i isolate A4/M1027 120M 7891 6748 129	<ul> <li>p): 100</li> <li>% identity A 98.49</li> <li>98.49</li> <li>98.49</li> <li>98.49</li> <li>98.49</li> <li>98.49</li> <li>98.49</li> </ul>	lignment length 465 465 465 465 465	7 7 7 7 7 7	0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867	19444 4782 19869 19181 35889	19908 extract	1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807		
Tankii Us 1 2 7 10 11 13 19 24	ng length (b) e TBLASTX [ A4/M1027 120M 7891 6748 129 139M S3131	p): 100 • • • • • • • • • • • • • • • • • •	lignment length 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 7	0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318	19444 4782 19869 19181 35889 36775 19090 4534	19908 extract - 5246 extract - 20333 extract - 19645 extract - 36353 extract - 37239 extract -	1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807		
lankii Us 1 2 7 10 11 13 19 24 30	ng length (by e TBLASTX [ A4/M1027 120M 7891 6748 129 139M S3131 S4355	p): 100 **********************************	lignment length 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 7 7	0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318 215673	19444 4782 19869 19181 35889 36775 19090 4534	19908 extract - 5246 extract - 20333 extract - 19645 extract - 36353 extract - 37239 extract - 19554 extract - 4998 extract -	1 1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807		
Tankii Us Us 1 2 7 10 11 13 19 24 30 31	ng length (b) e TBLASTX [ A4/M1027 120M 7891 6748 129 139M S3131 S4355 14	p): 100 <b>5 identity A</b> 98.49 98.49 98.49 98.49 98.49 98.49 98.49 98.49 98.49 98.49 98.49 98.49 98.49 98.49 98.49	lignment length 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 7 7 12 7 7 7	0 0 0 0 0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318 215673 8	19444 4782 19869 19181 35889 36775 19090 4534 1363524	19908 extract 5246 extract 20333 extract 19645 extract 36353 extract 19554 extract 19554 extract 1363988 extract 1363988 extract 20247 extract 20247 extract	1 1 1 1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807 785		
Tankii Us Us 1 2 7 10 11 13 19 24 30 31 34	ng length (b) e TBLASTX [ A4M1027 120M 7891 6748 129 139M S3131 S4355 14 10	<ul> <li>b): 100</li> <li>c)</li> <lic)< li=""> <li>c)</li> <li>c)</li> <li>c)</li>     &lt;</lic)<></ul>	lignment length 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 7 12 7 7 7	0 0 0 0 0 0 0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318 215673 8 182380	19444 4782 19869 19181 35889 36775 19090 4534 1363524 5559	19908 extract _ 5246 extract _ 20333 extract _ 19645 extract _ 36353 extract _ 37239 extract _ 19554 extract _ 4998 extract _ 1363988 extract _ 6023 extract _	1 1 1 1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807 785 807		
ate id 1 2 7 10 11 13 19 24 30 31 34 35 46	ng length (b) e TBLASTX [ A4/M1027 120M 7891 6748 129 139M S3131 S4355 14 10 20 26 255	IOD         IOD           I         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49	lignment length 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 12 7 7 7 7 7 7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318 215673 8 182380 182815 182880 183179	19444 4782 19869 19181 35889 36775 19090 4534 1363524 5559 19783 18879 4982	19908 extract _ 5246 extract _ 2033 extract _ 19645 extract _ 36353 extract _ 37239 extract _ 19554 extract _ 19554 extract _ 1363988 extract _ 20247 extract _ 19343 extract _ 5446 extract _	1 1 1 1 1 1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 785 807 807 807		
ate id 1 2 7 10 11 13 19 24 30 31 34 35 46 52	ng length (b) e TBLASTX [ A4/M1027 120M 7891 6748 129 139M S3131 S4355 14 10 20 26 255 243	D)         100           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49	ignment length 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 12 7 7 7 7 7 7 7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318 215673 8 182380 182815 182880 183179 183381	19444 4782 19869 19181 35889 36775 19090 4534 1363524 5559 19783 19879 4982 4532	19908 settest = 5246 settest = 20333 settest = 19645 settest = 37239 settest = 4998 settest = 18554 settest = 18588 settest = 20247 settest = 5446 settest = 54496 settest =	1 J T T T T T T T T T T J J	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807 807		
late id 1 2 7 10 11 13 19 24 30 31 34 35 46 52 61	ng length (b) e TBLASTX [ A4M1027 120M 7891 6748 129 139M S3131 S4355 14 10 26 255 243 393	<ul> <li>p): 100</li> <li>identity A</li> <li>98.49</li> </ul>	Ignment length 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182044 182318 215673 8 182380 182815 182880 182815 182880 183179 183381 183648	19444 4782 19869 19181 35889 36775 19090 4534 1363524 5559 19783 18879 4982 4532 7217	19908 estract 5246 extract 19945 extract 36353 extract 37239 extract 19554 extract 4998 extract 20247 extract 20247 extract 19348 extract 4996 extract 4996 extract	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807 807		
Tankii Us Us Us Us Us Us Us Us Us Us Us Us Us	ng length (b) e TBLASTX [ A4/M1027 120M 7891 6748 129 139M S3131 S4355 14 10 20 26 255 243 393 254	p):         100           *5:         identity A           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49	ignment length 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 12 7 7 7 7 7 7 7 7 7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318 215673 8 182380 182815 182880 183815 183881	19444 4782 19869 19181 35889 36775 19090 4534 1363524 5559 19783 18879 4982 4532 7217 4716	19908 senset - 5246 senset - 2033 senset - 19945 senset - 37239 senset - 37239 senset - 19554 senset - 19638 senset - 106398 senset - 19349 senset - 4996 senset - 4966 senset - 7681 senset - 5180 senset - 5180 senset -	+++++++++++++++++++++++++++++++++++++++	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807 807		
Elankii Us Us Us Us Us Us Us Us Us Us Us Us Us	ng length (b) e TBLASTX [ A4M1027 120M 7891 6748 129 139M S3131 S4355 14 10 20 26 255 243 393 254 25611	p)         100           •         identity A           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49	ignment length 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 12 7 7 7 7 7 7 7 7 7 7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318 215673 8 182380 182815 182880 182815 182880 183179 183381 183648 183818 183818 183818	19444 4782 19869 19181 35889 36775 19090 4534 1363524 5559 19783 18879 4982 4532 7217 7217 4716 4804	19908 strat - 5246 strat - 19908 strat - 19965 strat - 3033 strat - 33723 strat - 34723 strat - 19554 strat - 4998 strat - 19554 strat - 19934 strat - 19543 strat - 19343 strat - 5446 strat - 7681 strat - 7681 strat - 7681 strat - 5180 strat - 5180 strat - 5180 strat - 5180 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 35888 strat -	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807 807		
Flankii Us late id 2 7 10 11 13 19 24 30 31 34 35 46 52 61 64 67 82	ng length (b) e TBLASTX [ A4/M1027 120M 7891 6748 129 139M S3131 S4355 14 10 20 26 255 243 393 254 S5611 11-004	p):         100           *         identity A           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49	Ignment lengtil 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7		180177 180583 180965 181186 181867 182004 182318 215673 8 182380 182815 182880 182815 182881 183818 183818 183483 1844297	19444 4782 19869 19181 35889 36775 19090 4534 1363524 5559 19783 18879 4982 4532 7217 4716 4804 19518	19908 strat	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807 807		
Flankii Us late id 2 7 10 11 13 19 24 30 31 34 35 46 52 61 64 67	ng length (b) e TBLASTX [ A4M1027 120M 7891 6748 129 139M S3131 S4355 14 10 20 26 255 243 393 254 25611	p)         100           •         identity A           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49           98.49         98.49	ignment length 465 465 465 465 465 465 465 465 465 465	7 7 7 7 7 7 7 12 7 7 7 7 7 7 7 7 7 7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	180177 180583 180965 181186 181867 182004 182318 215673 8 182380 182815 182880 182815 182880 183179 183381 183648 183818 183818 183818	19444 4782 19869 19181 35889 36775 19090 4534 1363524 5559 19783 18879 4982 4532 7217 7217 4716 4804	19908 strat - 5246 strat - 19908 strat - 19965 strat - 3033 strat - 33723 strat - 34723 strat - 19554 strat - 4998 strat - 19554 strat - 19934 strat - 19543 strat - 19343 strat - 5446 strat - 7681 strat - 7681 strat - 7681 strat - 5180 strat - 5180 strat - 5180 strat - 5180 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 3588 strat - 35888 strat -	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	807 807 807 807 807 807 807 807 807 807		

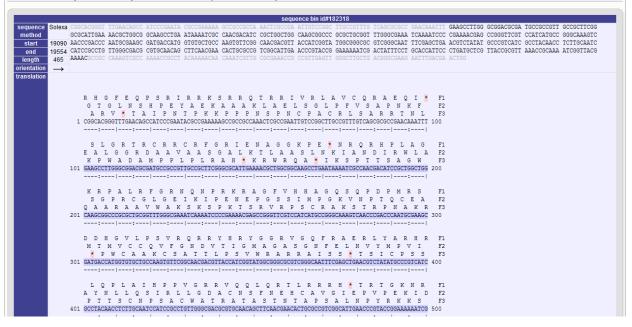
Clicking any of the 'extract' buttons will display the matched sequence along with a translated sequence and flanking sequences.

181867	35889	36353 extract →	$\rightarrow$	0.0	807
182004	36775	37239 extract →	$\rightarrow$	0.0	807
182318	19090	19554 extract ->	$\rightarrow$	0.0	807
215673	4534	4998 extract →	←	0.0	807
8	1363524	1363988 extract ->	$\rightarrow$	0.0	785
182380	5559	6023 extract →	$\rightarrow$	0.0	807
182815	19783	20247 extract →	$\rightarrow$	0.0	807

Query: Search | Browse | Profile/ST | List Breakdown: Isolate fields | Scheme/alleles | Publications

Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Extracted sequence: Seqbin id#:182318 (19090-19554)



At the bottom of the results table are links to export the matching sequences in FASTA format, (optionall) including flanking sequences. You can also export the table in tab-delimited text or Excel formats.

	00011	30.43	400	1	v	104143	4004	JZ00 extract ->	<del></del>	0.0	007
82	11-004	98.49	465	7	0	184297	19518	19982 extract →	$\rightarrow$	0.0	807
84	IAL2229	98.49	465	7	0	184635	4530	4994 extract -	←	0.0	807
90	CN100	98.49	465	7	0	184814	5350	5814 extract -	←	0.0	807

### 12.7.1 Include in results table fieldset

This selection box allows you to choose which isolate provenance fields will be included in the results table.

<ul> <li>Include in results tabl</li> </ul>	e —
country	
region	
year	-
epidemiological year	
age yr	
age mth	
sex	
disease	
source	
epidemiology	-

Multiple values can be selected by clicking while holding down Ctrl.

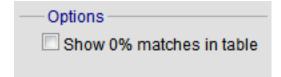
#### 12.7.2 Parameters fieldset

This section allows you to modify BLAST parameters. This affects sensitivity and speed.

-Parameters		
BLASTN word size:	11 🗸 👔	
BLASTN scoring:	reward:2; penalty:-3; gap open:5; gap extend:2 🚽	i
Hits per isolate:	1 👻	
Flanking length (bp):	100 👻 🚺	
🔲 Use TBLASTX 👔		

- BLASTN word size This is the length of the initial identical match that BLAST requires before extending a match (default: 11). Increasing this value improves speed at the expense of sensitivity.
- BLASTN scoring This is a dropdown box of combinations of identical base rewards; mismatch penalties; and gap open and extension penalties. BLASTN has a constrained list of allowed values which reflects the available options in the list.
- Hits per isolate By default, only the best match is shown. Increase this value to the number of hits you'd like to see per isolate.
- Flanking length Set the size of the upstream and downstream flanking sequences that you'd like to include.
- Use TBLASTX This compares the six-frame translation of your nucleotide query sequence against the six-frame translation of the contig sequences. This is significantly slower than using BLASTN.

#### 12.7.3 No matches



Click this option to create a row in the table indicating that a match was not found. This can be useful when screening a large number of isolates.

#### 12.7.4 Filter fieldset

This section allows you to further filter your collection of isolates and the contig sequences to include.

-Restrict included sec	quences by	
Sequence method:	<b>▼</b> [i	
Project:		<b>▼</b> [i
Experiment:	▼ i	

Available options are:

- Sequence method Choose to only analyse contigs that have been generated using a particular method. This depends on the method being set when the contigs were uploaded.
- Project Only include isolates belonging to the chosen project. This enables you to select all isolates and filter to a project.
- Experiment Contig files can belong to an experiment. How this is used can vary between databases, but this enables you to only include contigs from a particular experiment.

# **12.8 BURST**

BURST is an algorithm used to group MLST-type data based on a count of the number of profiles that match each other at specified numbers of loci. The analysis is available for both sequence definition database and isolate database schemes that have primary key fields set. The algorithm has to be *specifically enabled* by an administrator. Analysis is limited to 1000 or fewer records.

The plugin can be accessed following a query by clicking the 'BURST' button in the Analysis list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

	7891	DEA: TAOEA	UK		invasive (unspecified/other)		W	1101	ST-22 complex	-		50.4
		B54; Z1054	Finland		invasive (unspecified/other)		A	5	ST-5 complex/subgroup III	20	9	F3-1
9 C	00242007		UK		invasive (unspecified/other)		В	1102	ST-18 complex		14	
	0021/84				invasive (unspecified/other)		W	114	ST-22 complex		_	
0		B73; Z1073	Canada		invasive (unspecified/other)		A	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
1	129	B92; Z1092	Germany		invasive (unspecified/other)		Α	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
	0090/89				invasive (unspecified/other)		в	1015	ST-32 complex/ET-5 complex	7	16	
3	139M	B99; Z1099	Philippines	1968		Neisseria meningitidis	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
	0120/95					Neisseria meningitidis	Х	117			14	
5	1		Germany	1999		Neisseria meningitidis	E	864				
6	2		Germany	1999	carrier	Neisseria meningitidis	В	854	ST-18 complex			
7	3		Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
8	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
9	S3131	B213; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	Α	4	ST-4 complex/subgroup IV	7	13-1	F1-5
0	5		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
1	6		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
2	7		Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
3	8		Germany	1999	carrier	Neisseria meningitidis	В	32	ST-32 complex/ET-5 complex			
4	S4355	B227; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex/subgroup III	5-1	9	F3-1
5	9		Germany	1999	carrier	Neisseria meningitidis	в	930	ST-334 complex			

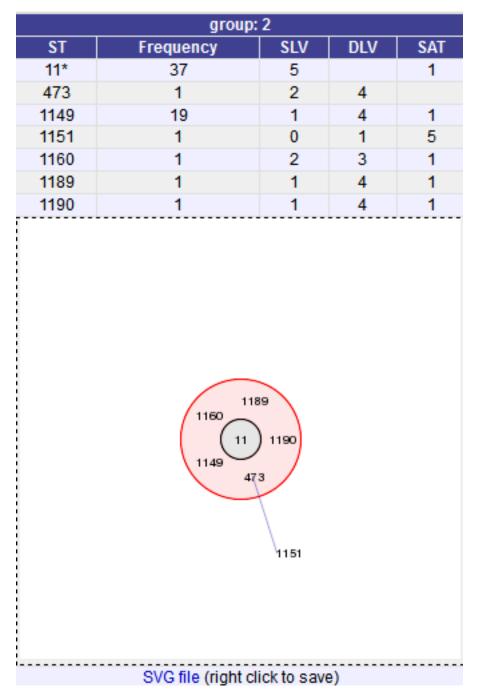
If there multiple schemes that can be analysed, these can then be selected along with the group definition.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/Sequences definitions   Database submissions	
BURST analysis	
Only of a strength MI OT	equence type (ST). bble.

Modifying the group definition affects the size of groups and how they link together. By default, the definition is n-2 (where n is the number of loci), so for example on a 7 locus MLST scheme groups contain STs that match at 5 or more loci to any other member of the group.

#### Click Submit.

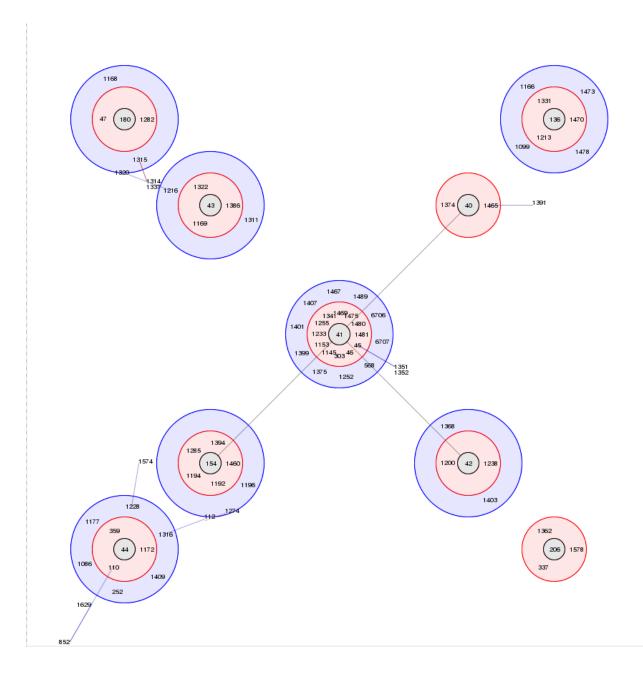
A series of tables will be displayed indicating the groups of profiles. Where one profile can be identified as a central genotype, i.e. the profile that has the greatest number of other profiles that are single locus variants (SLV), double locus variants (DLV) and so on, a graphical representation will be displayed. The central profile is indicated with an asterisk.



SLV profiles that match the central profile are shown within a red circle surrounding the central profile. Most distant profiles (triple locus variants) may be linked with a line. Larger groups may additionally have DLV profiles. These are shown in a blue circle.

group: 6									
ST	Frequency	SLV	DLV	SAT					
32*	2	3	2						
230	1	1	3	1					
484	1	0	3	2					
1015	1	1	4						
1100	1	1	2	2					
1148	1	0	4	1					
	SVG file (right cli	ck to save	e)						

Groups can get very large, where linked profiles form sub-groups and an attempt is made to depict these.



# 12.9 Codon usage

The codon usage plugin for isolate databases calculates the absolute and relative synonymous codon usage by isolate and by locus.

The function can be selected by clicking the 'Codon usage' link in the Analysis section of the main contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publication Links: Contents   Home   Options   Profiles/sequences of	
	hat represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition late may be submitted to this database and consequently it should be noted that it does not represent a population sample.
Query database - Search database - advanced queries. Browse database - peruse all records. - Search by combinations of loci (profiles) - including partial matchin - List query - find isolates by matching a field to an entered list. - Projects - main projects defined in database.	<ul> <li>Option settings         <ul> <li>Set general options - including isolate table field handling.</li> <li>Set display and query options for locus, schemes or scheme fields.</li> </ul> </li> <li>Ig.         <ul> <li>Set display and query options for locus, schemes or scheme fields.</li> <li>General information</li> <li>Isolates: 34221</li> <li>Last updated: 2015-07-02</li> <li>Update history</li> <li>About BIGSdb</li> </ul> </li> </ul>
Breakdown       Export         • Single field       • Export dataset         • Unique combinations       • Contigs         • Scheme and alleles       • Sequences - XMFA / concatenate         • Publications       • Sequence bin	ed FASTA formats BLAST

Alternatively, it can be accessed following a query by clicking the 'Codons' button in the Analysis list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

000242007         Czee           0021/84         Czee           0         6748           1         129           992; Z1092         C           2         0090/89           3         139M           4         0120/95	UK         2000           Finland         1975           UK         2000           ch Republic         1984           Canada         1971           Germany         1964           ch Republic         1989           hilippines         1968	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W A B W A A	1101 5 1102 114 1 1	ST-32 complex/ET-5 complex ST-22 complex ST-5 complex/subgroup III ST-18 complex ST-22 complex ST-1 complex/subgroup //II ST-1 complex/subgroup //II ST-32 complex/ET-5 complex	7 20 18-1 5-2	16 9 14 3 10	F3-1 F5-1 F3-6
7891         854; Z1054           M00242007         Czee           0021184         Czee           06748         B73; Z1073           129         B92; Z1092         C           2009089         Czee           3139M         B99; Z109         C           4012095         Czee         C	Finland         1975           UK         2000           ch Republic         1984           Canada         1971           Germany         1964           ch Republic         1989           hilippines         1968	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B W A A	5 1102 114 1 1	ST-5 complex/subgroup III ST-18 complex ST-22 complex ST-1 complex/subgroup I/II ST-1 complex/subgroup I/II	18-1 5-2	14 3 10	F5-1
M00242007         Czee           0021/84         Czee           0         6748         B73; Z1073           1         129         B92; Z1092         C           2         0090/89         Czee           3         139M         B99; Z1099         PI           4         0120/95         Czee	UK 2000 ch Republic 1984 Canada 1971 Germany 1964 ch Republic 1989 hilippines 1968	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B W A A	1102 114 1 1	ST-18 complex ST-22 complex ST-1 complex/subgroup I/II ST-1 complex/subgroup I/II	18-1 5-2	14 3 10	F5-1
0021/84         Czer           0         6748         B73; Z1073         0           1         129         B92; Z1092         C           2         0090/89         Czer         Czer           3         139M         B99; Z1099         PI           4         0120/95         Czer         Czer	ch Republic 1984 Canada 1971 Germany 1964 ch Republic 1989 hilippines 1968	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W A A	114 1 1	ST-22 complex ST-1 complex/subgroup I/II ST-1 complex/subgroup I/II	5-2	3 10	
0         6748         B73; Z1073         0           1         129         B92; Z1092         0           2         0090/89         Czer           3         139M         B99; Z1099         PI           4         0120/95         Czer         Czer	Canada 1971 Germany 1964 ch Republic 1989 hilippines 1968	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A	1	ST-1 complex/subgroup I/II ST-1 complex/subgroup I/II	5-2	10	
1         129         B92; Z1092         CC           2         0090/89         Cze         Cze           3         139M         B99; Z1099         PI           4         0120/95         Cze         Cze	Germany 1964 ch Republic 1989 hilippines 1968	invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A	1	ST-1 complex/subgroup I/II	5-2	10	
2 0090/89 Cze 3 139M B99; Z1099 Pr 4 0120/95 Cze	ch Republic 1989 hilippines 1968	invasive (unspecified/other)	Neisseria meningitidis						F3-6
3 139M B99; Z1099 Pt 4 0120/95 Cze	hilippines 1968	· · · · · · · · · · · · · · · · · · ·	~	B	1015	ST-32 complex/ET-5 complex	-		
4 0120/95 Cze			Main a sala sa sala sitisfa				7	16	
	ch Republic 1995		Neisseria meningitidis	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
5 1 0		invasive (unspecified/other)	Neisseria meningitidis	Х	117			14	
	Germany 1999	carrier	Neisseria meningitidis	E	864				
6 2 0	Germany 1999	carrier	Neisseria meningitidis	В	854	ST-18 complex			
7 3 0	Germany 1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
B 4 (	Germany 1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
9 \$3131 B213; Z1213	Ghana 1973	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex/subgroup IV	7	13-1	F1-5
0 5 0	Germany 1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
	Germany 1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
2 7 (	Germany 1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
3 8 0	Germany 1999	carrier	Neisseria meningitidis	В	32	ST-32 complex/ET-5 complex			
4 \$4355 B227; Z1227 E	Denmark 1974	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex/subgroup III	5-1	9	F3-1
5 9 0	Germany 1999	carrier	Neisseria meningitidis	в	930	ST-334 complex			

Enter the ids of the isolate records to analyse - these will be already entered if you accessed the plugin following a query. Select the loci you would like to analyse, either from the dropdown loci list, and/or by selecting one or more schemes.

Output         Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Public           Links: Contents   Home   Options   Profiles/sequent			
			Toggle: i
Codon usage analysis			
		database containing sequences, or with sequences tagged, can be inclu ing frame 1). Partial sequnces from the sequence bin will not be analysed	
	ier — Loci —	Schemes	
Paste in list of ids to include, start a new isolate include all ids. country region year age of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of the start of	aceF (NEIS1279) ackA2 (NEIS1727) achA (NEIS1729) • All None Pastelist	Control Control Control Proce     Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Contro Control Control Control Control	
Sequence retrieval	Codons	Action	
If both allele designations and tagged sequences exist for a locus, choose how you want these handled: [] Use sequences tagged from the bin Use allele sequence retrieved from external database	Select codon order:	Submit	
Ose allele sequences with problem flagged (defined alleles w	vill still be used)		

Click submit. The job will be submitted to the queue and will start running shortly. Click the link to follow the job progress and view the output.

Query: Search  Browse   Profile/ST  List Breakdown: Isolate fields   Scheme/allels   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
	Toggle: i
Codon usage analysis	
This analysis has been submitted to the job queue.	
Please be aware that this job may take some time depending on the number of sequences to analyse and how busy the server is.	
Follow the progress of this job and view the output.	

Four tab-delimited text files will be created.

- Absolute frequency of codon usage by isolate
- Absolute frequency of codon usage by locus
- Relative synonymous codon usage by isolate
- Relative synonymous codon usage by locus

PtrisMJ_ST Breakdown: Isolate fields   Scheme/alleles   P Links: Contents   Home   Options   Profiles/se	
Job status viewer	
Status  Job id: BiGSdb_13269_1405586315_76138 Submit time: 2014-07-17 09:38:35 Status: finished Start time: 2014-07-17 09:38:39 Progress: toos. Stop time: 2014-07-17 09:38:05 Total time: 26 seconds  Output  Absolute frequency of codon usage by isolate Absolute frequency of codon usage by isolate Absolute frequency of codon usage by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate Relative synonymous codon usage (RSCU) by isolate	
Please note that job results will remain on the server for 7 days.	

# 12.10 Unique combinations

The Unique Combinations plugin calculates the frequencies of unique file combinations within an isolate dataset. Provenance fields, composite fields, allele designations and scheme fields can be combined.

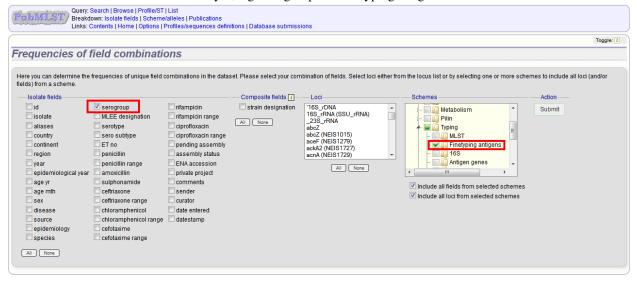
The function can be selected by clicking the 'Unique combinations' link in the Breakdown section of the main contents page. This will run the analysis on the entire database.

Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions									
	epresent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition may be submitted to this database and consequently it should be noted that it does not represent a population sample. Option settings • Set general options - including isolate table field handling. • Set display and query options for locus, schemes or scheme fields. • Isolates: 34221 • Last updated: 2015-07-02 • Update history • About BIGSdb								
Breakdown  Single field  Won field  Unique combinations  Strieme and alteres  Publications  Sequence bin  Export  Export  Export dataset  Contigs  Sequences - XMFA / concatenated Field  Sequences - XMFA / concatenated Field  Contigs  Sequence bin  Export   Analysis Codon usage Presence/absence status of loci Genome comparator BLAST BLAST Miscellaneous Description of database fields									

Alternatively, it can be accessed following a query by clicking the 'Combinations' button in the Breakdown list at the bottom of the results table. This will run the analysis on the dataset returned from the query. Please note that the list of functions here may vary depending on the setup of the database.

5 M0024 6 M0028 7 789 8 M0024 9 0021 10 674 11 12	82207 891 B54; Z1054 42007	UK UK Finland		invasive (unspecified/other)					7	10	
7 789 8 M0024 9 0021 10 674	891 B54; Z1054 42007		2000		~	В		ST-32 complex/ET-5 complex	(	16	
8 M0024 9 0021 10 674	42007			invasive (unspecified/other)		W	1101	ST-22 complex			
9 0021 10 674				invasive (unspecified/other)		A	5	ST-5 complex/subgroup III	20	9	F3-1
10 674		UK		invasive (unspecified/other)		В	1102	ST-18 complex		14	
				invasive (unspecified/other)	~	W	114	ST-22 complex			
11 12		Canada		invasive (unspecified/other)	-	Α	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
		Germany		invasive (unspecified/other)		A	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
0090				invasive (unspecified/other)		В	1015	ST-32 complex/ET-5 complex	7	16	
13 139			1968		Neisseria meningitidis	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
14 0120	0/95			invasive (unspecified/other)	Neisseria meningitidis	Х	117			14	
15 1			1999		Neisseria meningitidis	E	864				
16 2	2	Germany	1999	carrier	Neisseria meningitidis	в	854	ST-18 complex			
17 3	3	Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
18 4	4	Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
19 S31	131 B213; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	Α	4	ST-4 complex/subgroup IV	7	13-1	F1-5
20 5	5	Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
21 6		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
22 7	7	Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
23 8	8	Germany	1999	carrier	Neisseria meningitidis	В	32	ST-32 complex/ET-5 complex			
24 S43	355 B227; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex/subgroup III	5-1	9	F3-1
25 9	9	Germany	1999	carrier	Neisseria meningitidis	В	930	ST-334 complex			

Select the combination of fields to analyse, e.g. serogroup and finetyping antigens.



Click submit. When the analysis has completed you will see a table showing the unique combinations of the selected fields along with the frequency and percentage of the combination.

						Toggle
quen	cies of f	ïeld coi	mbinat	ions		
mber of uni	ique combinati	ons: 2939				
e percentag	jes may add up	p to more thar	n 100% if you	have selected	d loci or scheme	fields with multiple values for an isolate.
rogroup 🖨	PorA VR1 🖨	Por∆ VR2 ≜	Fet∆ VR ♠	Frequency 🖨	Percentage 🗢	
B	7-2	4	F1-5	625	5.65	
w	5	2	F1-1	513	4.64	
В	19	15	F5-1	385	3.48	
С	5	2	F3-6	283	2.56	
-	18-10	43	F3-14	281	2.54	
в	7	16	F3-3	252	2.28	
С	5-1	10-8	F3-6	222	2.01	
В	22	9	F5-12	171	1.55	
в	22	14	F5-5	167	1.51	
Α	5-2	10	F3-5	135	1.22	
в	19-1	15-11	F5-1	107	0.97	
Y	5-1	10-1	F4-1	96	0.87	
	5	2	F1-1	93	0.84	
Y	5-2	10-1	F4-1	79	0.71	
С	5	2	F5-8	79	0.71	
W	18-1	3	F4-1	77	0.70	
В	22-1	14	F4-1	77	0.70	
С	5	2	F3-3	76	0.69	
В	18-1	3	F1-5	72	0.65	
Y	5-1	2-2	F5-8	69	0.62	
В	19	15	F1-14	61	0.55	
B	7-2	13-2 16	F1-5 F3-3	61 56	0.55	
в	7-2	16	F3-3	36	0.51	

The table can be downloaded in tab-delimited text or Excel formats by clicking the links at the bottom of the page.

B	19	13-2	FF 4		
D		10 2	F5-1	1	0.01
D	12.1	13.1	F3-29	1	0.01
NG	7-2	16-126	F3-3	1	0.01
в	7-2	30	F5-1	1	0.01
в	7	30-3	F1-14	1	0.01
в	5-1	10-10	F3-16	1	0.01
	7-2	30-3	F1-7	1	0.01
С	5-2	10	F1-7	1	0.01
NG	18-1	3	F1-34	1	0.01
NG	22	14-6	F4-2	1	0.01
в	7-2	4	F1-88	1	0.01
NG	5-1	10-62	F1-3	1	0.01
Y	12-3	4	F4-1	1	0.01
NG	7-2	16	F1-104	1	0.01
Х	12-1	16-52	F3-9	1	0.01
NG	18-4	35-34	F4-1	1	0.01
	oad as tab-de oad in Excel fo				

# 12.11 Polymorphisms

The Polymorphisms plugin generates a *Locus Explorer* polymorphic site analysis on the alleles designated in an isolate dataset following a query.

The analysis is accessed by clicking the 'Polymorphic sites' button in the Breakdown list at the bottom of a results table following a query.

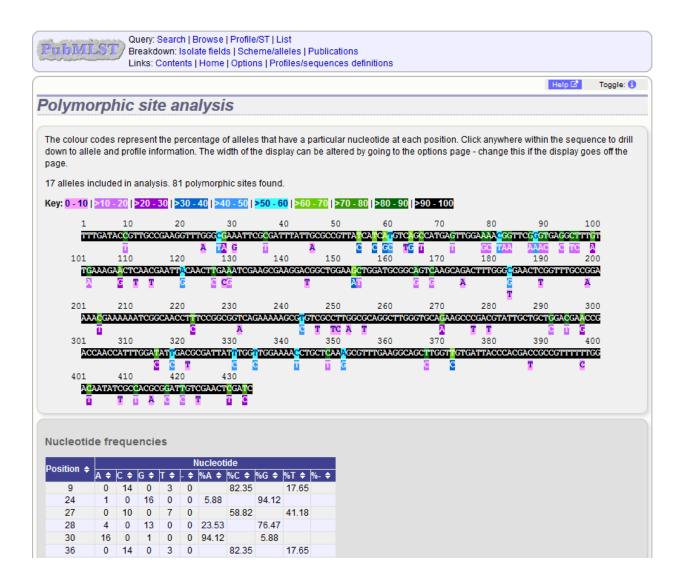
						-						
10	6748	B73; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	Α	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
11	129	B92; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	Α	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
12	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningitidis	в	1015	ST-32 complex/ET-5 complex	7	16	
13	139M	B99; Z1099	Philippines	1968		Neisseria meningitidis	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
14	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitidis	Х	117			14	
15	1		Germany	1999	carrier	Neisseria meningitidis	E	864				
16	2		Germany	1999	carrier	Neisseria meningitidis	в	854	ST-18 complex			
17	3		Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
19	S3131	B213; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	Α	4	ST-4 complex/subgroup IV	7	13-1	F1-5
20	5		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
21	6		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
22	7		Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningitidis	в	32	ST-32 complex/ET-5 complex			
24	S4355	B227; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex/subgroup III	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria meningitidis	в	930	ST-334 complex			
Analysis tools: Breakdown: Fields Two Field Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status Analysis: EURBT   Codons   Presence/Absence   Genome Comparator   BLAST												
Page		3 4 5	igs Sequences		Last							

Select the locus that you would like to analyse from the list.

Problem Search   Brows   Profile/ST  List Breakown: Solate fields   Schemelaileles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Polymorphic site analysis	Toggle: i
This tool will analyse the polymorphic sites in the selected locus for the current isolate dataset. If more than 50 sequences have been selected, the job will be run by the offline job manager which may take a few minutes (or longer depending on the queue). This is because sequences may have gaps in them and consequently need to be aligned which is a processor and memory-intensive operation. Loci Options Action 165 rRNA (SSU rRNA) 2.23 rRNA 2.23 rRNA 2.23 rRNA 2.23 rRNA 2.23 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2.25 rRNA 2	1

Click 'Analyse'.

A schematic of the locus is generated showing the polymorphic sites. A full description of this can be found in the *Locus Explorer polymorphic site analysis* section.



# 12.12 Presence/absence

This plugin displays the status of loci for isolate records. It will shown whether a locus has been designated with an allele name, has a sequence tag, or both.

The function can be selected by clicking the 'Presence/absence status of loci' link in the 'Analysis' section of the main contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in th database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does	
<ul> <li>Query database</li> <li>Search database - advanced queries.</li> <li>Browse database - peruse all records.</li> <li>Search by combinations of loci (pofiles) - including partial matching.</li> <li>List query - find isolates by matching a field to an entered list.</li> <li>Projects - main projects defined in database.</li> </ul>	Submissions Manage submissions General information Isolates: 34221 Last updated: 2015-07-02 Update history About BIGSdb
Breakdown       Export       Analysis       Miscellaneo         • Single field       • Export dataset       • Contigs       • Contigs       • Description of         • Unique combinations       • Scheme and alleles       • Sequences - XMFA / concatenated FASTA formats       • Codon usage       • Description of         • Publications       • Sequence bin       • Sequence - XMFA / concatenated FASTA formats       • BLAST       • Description of	ous of database fields

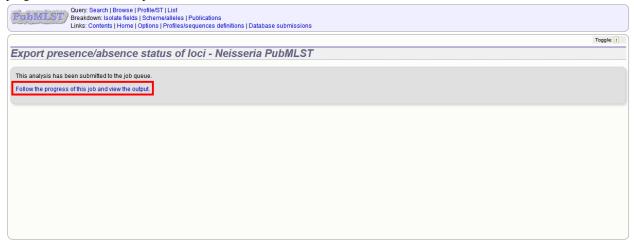
Alternatively, it can be accessed following a query by clicking the 'Presence/Absence' button in the Analysis list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

5	M00240227		UK	2000	invasive (unspecified/other)	Neisseria meningitid	is B	1100	ST-32 complex/ET-5 complex	7	16	
6	M00282207		UK	2000	invasive (unspecified/other)	Neisseria meningitid	is W	1101	ST-22 complex			
7	7891	B54; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitid	is A	5	ST-5 complex/subgroup III	20	9	F3-1
8	M00242007		UK	2000	invasive (unspecified/other)	Neisseria meningitid	is B	1102	ST-18 complex		14	
9	0021/84		Czech Republic	1984	invasive (unspecified/other)	Neisseria meningitid	is W	114	ST-22 complex			
10	6748	B73; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitid	is A	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
11	129	B92; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitid	is A	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
12	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningitid	is B	1015	ST-32 complex/ET-5 complex	7	16	
13	139M	B99; Z1099	Philippines	1968		Neisseria meningitid	is A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
14	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitid	is X	117			14	
15	1		Germany	1999	carrier	Neisseria meningitid	is E	864				
16	2		Germany	1999	carrier	Neisseria meningitid	is B	854	ST-18 complex			
17	3		Germany	1999	carrier	Neisseria meningitid	is W	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningitid	is B	19	ST-18 complex			
19	S3131	B213; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitid	is A	4	ST-4 complex/subgroup IV	7	13-1	F1-5
20	5		Germany	1999	carrier	Neisseria meningitid	is NG	198	ST-198 complex			
21	6		Germany	1999	carrier	Neisseria meningitid	is NG	198	ST-198 complex			
22	7		Germany	1999	carrier	Neisseria meningitid	is E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningitid	is B	32	ST-32 complex/ET-5 complex			
24	S4355	B227; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitid	is A	5	ST-5 complex/subgroup III	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria meningitid	is B	930	ST-334 complex			
Ana	lysis tools	:										
		elds Two Fi JRST Codo			Combinations Schemes/	AST Publications	Sequence	bin	Tag status			
	nalysis: BL		ns Presence/Al				Sequence	bin	Tag status			

Enter the ids of the isolate records to analyse - these will be already entered if you accessed the plugin following a query. Select the loci you would like to analyse, either from the dropdown loci list, and/or by selecting one or more schemes.

Query: Search   Browse   Pro Breakdown: Isolate fields   S Links: Contents   Home   Opt	cheme/alleles   Publications ions   Profiles/sequences defini				Toggle: 1
This script will export data showing whether a loca are members of the scheme.			eck the loci that you would like to include. Alte	rnatively select one or more schemes to	o include all loci that
Select ids Paste in list of ids to include, start a new line for each. Leave blank to include all ids. 1 2 3 4 5 6	Include in identifier isolate country region epidemiological year age yr age yr sex disease source	Loci 165 _FRA (SSU_FRNA) 235 _FRNA abc2 (NEIS1015) ac#A (NEIS1729) ackA2 (NEIS1729) ackA2 (NEIS1729) All None Paste lat	Schemes Genetic Information Proce Metabolism Pilin Typing Typing Theyping antigens 165 165	Options Mark present if : e either designations of tags set allele designations defined sequence tags defined Symbol for present O - Symbol for absent X - Generate distance matrix []	Action

Click submit. The job will be submitted to the queue and will start running shortly. Click the link to follow the job progress and view the output.



When complete, a single text file will have been generated.

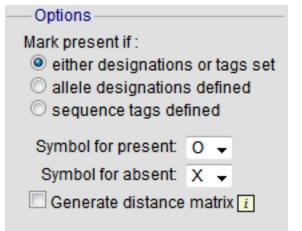
PriDIALST Breakdown: Isolate Fields   Schemeralicities   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Job status viewer
Status
Job id:         BIGSdb_31442_1405591304_69061           Submit time:         2014407-17 11:0144           Status:         finished           Start time:         2014-07-17 11:01:44           Progress:         60%           Stop time:         2014-07-17 11:01:49           Total time:         2014-07-17 11:01:49
Output  Main output file
Please note that job results will remain on the server for 7 days.

This is a tab-delimited text file that uses 'O' to represent presence and 'X' to represent a missing locus designation or tag.

id	pgm	adk	abcZ	pdhC	gdh	fumC	aroE
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0

#### 12.12.1 Options

There are a number of options that can be selected to modify the output.



With these you can change the symbols used and whether designations, or tags, or both are counted.

You can also choose to generate a distance matrix based on presence/absence.

# 12.13 Tag status

The tag status plugin displays a graphical representation of the status of loci designations or tags for isolate data. It is accessed following a query by clicking the 'Tag status' button in the Breakdown section at the bottom of the results table.

id	ate prove		nenot ▼ <	ype fields —	<b>-</b> 10		+	Display/sort option Order by: id Display: 25	records	s per p	▼ ascending age i	g 🔻			Modify form options
Actio	n —														
Rese	et S	ubmit													
9 records	returned	J. CIICK II	e nyp	erlinks for de	tailed ii	nformation.									
					l	solate fields 👔					MLST	Fine	typing ant	igens	
id isc	olate	aliase	;	country	year	disease	;	species	serogroup	ST	cional complex	PorA VR1	PorA VR	2 FetA VR	
1 A4/N	/1027	B1; Z10	1	USA	1937	invasive (unspeci	fied/other)	Neisseria meningitidis	A	4	ST-4 complex/subgroup IV	5-2	10	F1-5	
		B35; Z10	35	Pakistan	1967	meningitis and se	epticaemia	Neisseria meningitidis	Α	1	ST-1 complex/subgroup I/II	5-2	10	F5-1	
3 M002				UK				Neisseria meningitidis	в	1099		19	15		
4 M1	1027 E	B43; Z10	43	USA	1937	invasive (unspeci	fied/other)	Neisseria meningitidis	A	4	ST-4 complex/subgroup IV				
5 M002				UK				Neisseria meningitidis	В		ST-32 complex/ET-5 complex	7	16		
6 M002				UK				Neisseria meningitidis	W	1101	ST-22 complex				
		B54; Z10	54	Finland				Neisseria meningitidis		5	ST-5 complex/subgroup III	20	9	F3-1	
8 M002				UK				Neisseria meningitidis	В	1102	ST-18 complex		14		
9 002	21/84		С	zech Republic	c 1984	invasive (unspeci	fied/other)	Neisseria meningitidis	W	114	ST-22 complex				
Analysi	is tools														
Analysi	13 10013														
Breakdo	wn: Fi	elds ]	wo Fi	eld Polymo	rphic sit	es Combinations	Scheme	s/alleles Publications	Sequence	bin )	Tag status				
Analy	sis: BL	JRST	Codo	ns Preseno	e/Absen	ce Genome Com	parator	BLAST							
	_														
Exp		ataset	Conti	gs Sequen	oes										

#### Select the loci you would like to analyse.

Print Browse   Profile/ST   List Breakdown: Isolate fields   Schemericalleise   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
	Toggle: i
Tag status	
Select schemes or groups of schemes within the tree. A breakdown of the individual loci belonging to these schemes will then be performed.	
Image: Select	

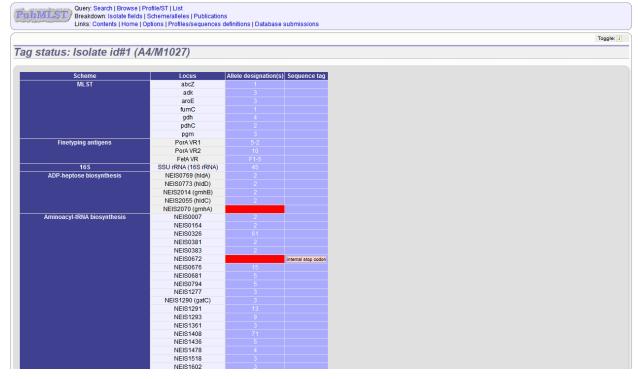
You should see a series of bars representing loci. The colour of these bars designates whether they have an allele designation only, a sequence tag only, both designations or tags, or whether they have flags set.

									Toggle:
ag status									
Select schemes or groups of sch	nemes within the tree. /	A breakdown of the indivi	idual loci belonging	to these schemes will	hen be performed.				
Capsule Reg	gion A - Serogroup A gion A - Serogroup B gion A - Serogroup C	A V F							
ars represent loci by schemes ;	arranged in alphabetic	al order. If a locus appea	ars in more than on	e scheme it will appear	more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown for	r an isolate.
Key Allele designated only   Sequenc					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown foi	r an isolate.
Key Viele designated only   Sequenc 2463 loci selected: Id   Isolate					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown foi	r an isolate.
Cey Viele designated only   Sequence 2463 loci selected: Id Isolate 1 A4/M1027					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown for	r an isolate.
Cey Illele designated only   Sequence 463 loci selected: I Isolate A4/M1027 2 120M					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown for	r an isolate.
Cey Illele designated only   Sequence 463 loci selected: 1 A4/M1027 2 120M 3 M00242905					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown for	r an isolate.
Key Illele designated only   Sequence 463 loci selected: 1 A4/M1027 2 120M 3 M00242905 4 M1027					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown for	r an isolate.
Ilele designated only   Sequence           463 loci selected:           1         A4/M1027           2         120M           3         M00242905           4         M1027           5         M00240227					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown for	r an isolate.
Idel         Isolate         Isolate           1         Ad/M1027         Isolate         Isolate           2         120M         Isolate         Isolate           3         M00242905         Isolate         Isolate           4         M1027         Isolate         Isolate           5         M00240207         Isolate         Isolate						s graphic. Click on the	id hyperlink for a det	ailed breakdown for	ran isolate.
Aurent         Sequence           2463         Loci selected:           1         Ad/M1027           2         120M           3         M00242905           4         M1027           5         M00240202           6         M0028207           7         7891					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown for	ran isolate.
1         A4/M1027           2         120M           3         M00242905           4         M1027           5         M00240227           6         M00282207					more than once in thi	s graphic. Click on the	id hyperlink for a det	ailed breakdown for	r an isolate.

Hovering the mouse over the bars will indicate the scheme represented.

Note: Loci will be represented more than once if they are members of multiple selected schemes.

Clicking any of the isolate id hyperlinks navigates to a page that breaks down the exact status for all loci of that isolate.



There is a column each for allele designations and sequence tags. If an allele designation is defined, the allele identifier

is displayed. Cells shaded in blue show that the designation or tag is present, whereas red indicates thet they are absent.

# **Data export plugins**

# 13.1 Isolate record export

You can export the entire isolate recordset by clicking the 'Export dataset' link in the Export section of the main contents page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions
Neisseria PubMLST database The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.
Query database       Search database - advanced queries.         Browse database - peruse all records.       Search by combinations of loci (profiles) - including partial matching.         List query - find isolates by matching a field to an entered list.       Set display and query options for locus, schemes or scheme fields.         Projects - main projects defined in database.       Submissions       Isolates: 34218         Last updated: 2015-06-30       Update history         Vision Barlow       Vision Barlow
<ul> <li>Breakdown</li> <li>Single field</li> <li>Two field</li> <li>Unique combinations</li> <li>Scheme and aleles</li> <li>Publications</li> <li>Sequence bin</li> </ul>

Alternatively, you can export the recordsets of isolates returned from a database query by clicking the 'Dataset' button in the Export list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

				Is	olate fields 👔				MLST	Finet	yping antig	jens
id	isolate	aliases	country	year	disease	species	serogroup	ST	clonal complex	PorA VR1	PorA VR2	FetA VR
1	A4/M1027	B1; Z1001	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	Α	4	ST-4 complex/subgroup IV	5-2	10	F1-5
2	120M	B35; Z1035	Pakistan	1967	meningitis and septicaemia	Neisseria meningitidis	Α	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
3	M00242905		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	в	1099		19	15	
4	M1027	B43; Z1043	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex/subgroup IV			
5	M00240227		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	в	1100	ST-32 complex/ET-5 complex	7	16	
6	M00282207		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	W	1101	ST-22 complex			
7	7891	B54; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex/subgroup III	20	9	F3-1
8	M00242007		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	В	1102	ST-18 complex		14	
9	0021/84		Czech Republic	1984	invasive (unspecified/other)	Neisseria meningitidis	W	114	ST-22 complex			
10	6748	B73; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	Α	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
11	129	B92; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	Α	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
12	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningitidis	В	1015	ST-32 complex/ET-5 complex	7	16	
13	139M	B99; Z1099		1968		Neisseria meningitidis	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
14	0120/95				invasive (unspecified/other)	Neisseria meningitidis	Х	117			14	
15	1		Germany	1999	carrier	Neisseria meningitidis	E	864				
16	2		Germany	1999	carrier	Neisseria meningitidis	в	854	ST-18 complex			
17	3		Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
19		B213; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	Α	4	ST-4 complex/subgroup IV	7	13-1	F1-5
20	5		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
21	6			1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
22	7		Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningitidis	в	32	ST-32 complex/ET-5 complex			
24	S4355	B227; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex/subgroup III	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria meningitidis	в	930	ST-334 complex			

#### Select the isolate fields and schemes to include.

							Toggle: i
Export dataset	taset in tab-delimited text, su	itable for importing into a	spreadsheet. Select which	fields you would li	ke included. Select loci either	from the locus list or by selecting one or more schemes to	include all
loci (and/or fields) from a sc locate fields V id V isolate allases continent region year epidemiological year age yt age mth sex	heme.	rifampicin rifampicin range diprofloxacin pending assembly assembly status ENA accession private project comments sender durator	Composite fields [	References references PubMed id Full citation	Loci 165.rDNA 165.rRNA (SSU_rRNA) _235_rRNA abc2 abc2 (NEIS1015) aceF (NEIS1279) aceF (NEIS1279) aceA (NEIS1727) aceA (NEIS1727) aceA (NEIS1729) All None	Schemes Venetic information Proce Plin Plin Plin Finetyping antigens Antioen genes Vinclude all fields from selected schemes Vinclude all fields from selected schemes	
disease source epidemiology species	chloramphenicol     chloramphenicol range     cefotaxime     cefotaxime range	date entered	Options Duclude locus commo Export allele numbers Use one row per field Include isolate field in Export full allele desi	s I 1 row (used only wi	th 'one row' option) ed only with 'one row' option)	Molecular weights Action C Export protein molecular weights GTG/TTG at start codes for methionine	]

Click Submit.

You can then download the data in tab-delimited text or Excel formats.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/Sequences definitions   Database submissions	
Top	ggle: i
Export dataset	
Please wait for processing to finish (do not refresh page).	
Output files being generated done	
Download: Text file   Excel file (right-click to save)	

#### 13.1.1 Advanced options

Options	
Include locus common names	
Export allele numbers	
Use one row per field	
Include isolate field in row (used only with 'one row' option)	
Export full allele designation record (used only with 'one row' optic	on)

The options fieldset has the following options.

- Include locus common names any common name for the locus is displayed in parentheses following the primary name.
- Export allele numbers the allele designation is included for any locus included.
- Use one row per field this is an alternative output format where instead of each locus and field having a separate column, each field is export on a separate row.
- Include isolate field in row the name of the isolate is included as a separate column when exporting in 'one row per field' fomrmat.
- Export full allele designation record export sender, curator and datestamp information as separate rows when exporting allele designation data.

### 13.1.2 Molecular weight calculation

The plugin can also calculate the predicted molecular weight of the gene product of any allele designated in the dataset.

Molecular weights

Export protein molecular weights

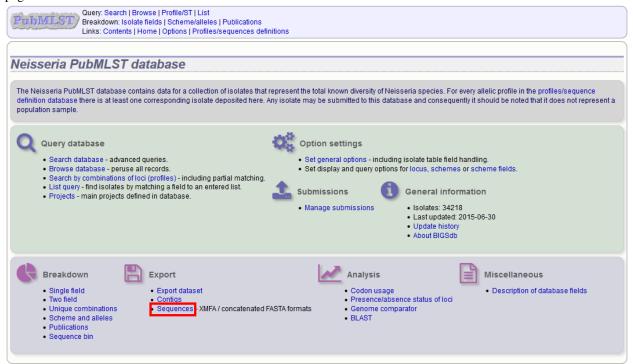
GTG/TTG at start codes for methionine

Click the 'Export protein molecular weight' checkbox. Additional columns (or rows depending on the output format) will be created to include the molecular weight data.

# 13.2 Sequence export

You can export the sequences for any set of loci designated in isolate records, or belonging to scheme profiles in the sequence definition database.

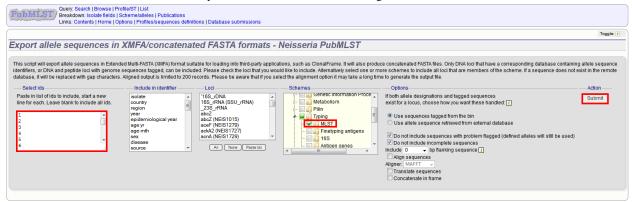
The sequence export function can be accessed by clicking the 'Sequences' link in the Export section of the contents page.



Alternatively, you can access this function by clicking the 'Sequences' button in the Export list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

873; Z1073 892; Z1092 899; Z1099	UK Finland UK Czech Republic Canada Germany	2000 1975 2000 1984 1971 1964	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B A B W A	1100 1101 5 1102 114	ST-32 complex/ET-5 complex ST-22 complex ST-5 complex/subgroup III ST-18 complex ST-22 complex	7 20	16 9 14	F3-1
873; Z1073 892; Z1092 899; Z1099	Finland UK Czech Republic Canada Germany	1975 2000 1984 1971 1964	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B W	5 1102	ST-5 complex/subgroup III ST-18 complex ST-22 complex	20		F3-1
873; Z1073 892; Z1092 899; Z1099	UK Czech Republic Canada Germany	2000 1984 1971 1964	invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B W	1102	ST-18 complex ST-22 complex	20		F3-1
873; Z1073 892; Z1092 899; Z1099	Czech Republic Canada Germany	1984 1971 1964	invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	W		ST-22 complex		14	
873; Z1073 892; Z1092 899; Z1099	Canada Germany	1971 1964	invasive (unspecified/other)	Neisseria meningitidis		114				
892; Z1092 899; Z1099	Germany	1964			A					
999; Z1099			invacive (unenacified/other)			1	ST-1 complex/subgroup I/II	18-1	3	F5-1
999; Z1099	Czech Republic		invasive (unspecifieu/outer)	Neisseria meningitidis	Α	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
		1989	invasive (unspecified/other)	Neisseria meningitidis	в	1015	ST-32 complex/ET-5 complex	7	16	
		1968		Neisseria meningitidis	Α	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
			invasive (unspecified/other)	Neisseria meningitidis	Х	117			14	
	Germany	1999	carrier	Neisseria meningitidis	E	864				
	Germany	1999	carrier	Neisseria meningitidis	в	854	ST-18 complex			
	Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
	Germany	1999	carrier	Neisseria meningitidis	в	19	ST-18 complex			
213; Z1213			invasive (unspecified/other)	Neisseria meningitidis	Α	4	ST-4 complex/subgroup IV	7	13-1	F1-5
	Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
	Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
	Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
	Germany	1999	carrier	Neisseria meningitidis	в	32	ST-32 complex/ET-5 complex			
227; Z1227			invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex/subgroup III	5-1	9	F3-1
	Germany	1999	carrier	Neisseria meningitidis	в	930	ST-334 complex			
	227; Z1227	Germany Germany Cermany Germany Germany Germany Germany Denmark Germany	Germany         1999           Germany         1999           Germany         1999           13, Z1213         Ghana           Germany         1999           2000         Germany           1999         1999           2000         1999           2000         1999           1999         1999           1999         1999           2000         1999           2000         1999           1999         1999           1990         1999           1990         1999           1990         1999           1999         1999           1999         1999           1999         1999           1999         1999           1999         1999           1999         1999           1999         1999           1999         1999 </td <td>Germany 1999 carrier Germany 1999 carrier 1973 invasive (unspecified)other) Germany 1999 carrier Germany 1999 carrier Germany 1999 carrier Germany 1999 carrier 227,21227 Demmark 1974 invasive (unspecified)other)</td> <td>Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis 113, Z1213 Ghana 1973 invashe (unspecifiediother) Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis</td> <td>Germany         1999         carrier         Nelsseria meningtidis         W           Germany         1999         carrier         Nelsseria meningtidis         B           1212 Z1213         Ghana         1973 invasive (unspecified/other) Neisseria meningtidis         A           Germany         1999         carrier         Neisseria meningtidis         NG           Germany         1999         carrier         Neisseria meningtidis         NG           Germany         1999         carrier         Neisseria meningtidis         E           Germany         1999         carrier         Neisseria meningtidis         E           Germany         1999         carrier         Neisseria meningtidis         B           227, Z1227         Demark         1974 invasive (unspecified/other) Neisseria meningtidis         B</td> <td>Germany         1999         carrier         Neisseria meningitidis         W         174           Germany         1999         carrier         Neisseria meningitidis         B         19           P13; Z1213         Ghana         1973         invasive (unspecified/detr)         Neisseria meningitidis         A         4           Germany         1999         carrier         Neisseria meningitidis         NG         198           227; Z1227         Demark         1974         invasive (unspecified/det)         NGisseria meningitidis         A         5</td> <td>Germany         1999         carrier         Nelsseria meninglitidis         W         174         67-174 complex           Germany         1999         carrier         Nelsseria meninglitidis         B         19         ST-18 complex           1213,21213         Gana         1973 invasive (unspecified/otter)         Nelsseria meninglitidis         A         ST-48 complex/budgroup IV           Germany         1999         carrier         Nelsseria meninglitidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meninglitidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meninglitidis         E         60         ST-60 complex           Germany         1999         carrier         Nelsseria meninglitidis         E         60         ST-60 complex           Germany         1999         carrier         Nelsseria meninglitidis         B         32         ST-32 complex/ST-32 complex           Germany         1999         carrier         Nelsseria meninglitidis         B         5         ST-50 complex/SU-32 complex           Cermany         1994         carrier         Nelsseria meninglitidis         B         5         <t< td=""><td>Germany         1999         carrier         Nelsseria meningtidis         W         174         ST-174 complex           Germany         1999         carrier         Nelsseria meningtidis         B         19         ST-18 complex           213,21213         Gramany         1973         invasive (unspecified/other)         Nelsseria meningtidis         A         4         ST-18 complex/subgroup IV         7           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         E         60         ST-60 complex           Germany         1999         carrier         Nelsseria meningtidis         B         32         ST-32 complex/EF-5 complex           Cermany         1999         carrier         Nelsseria meningtidis         B         32         ST-32 complex/EF-5 complex           227.21227         Demark         1974         invasibe (unspecified/duer)</td><td>Germany         1999         carrier         Neisseria meningtildis         W         174         ST-174 complex           Germany         1999         carrier         Neisseria meningtildis         B         19         ST-18 complex           P13, Z1213         Germany         1973         invasive (unspecified/other)         Neisseria meningtildis         A         ST-48 complex/subgroup IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         NG         198         ST-198 complex           Germany         1999         carrier         Neisseria meningtildis         NG         198         ST-198 complex           Germany         1999         carrier         Neisseria meningtildis         R         6         6         ST-60 complex           Germany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7</td></t<></td>	Germany 1999 carrier Germany 1999 carrier 1973 invasive (unspecified)other) Germany 1999 carrier Germany 1999 carrier Germany 1999 carrier Germany 1999 carrier 227,21227 Demmark 1974 invasive (unspecified)other)	Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis 113, Z1213 Ghana 1973 invashe (unspecifiediother) Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis Germany 1999 carrier Neisseria meningitudis	Germany         1999         carrier         Nelsseria meningtidis         W           Germany         1999         carrier         Nelsseria meningtidis         B           1212 Z1213         Ghana         1973 invasive (unspecified/other) Neisseria meningtidis         A           Germany         1999         carrier         Neisseria meningtidis         NG           Germany         1999         carrier         Neisseria meningtidis         NG           Germany         1999         carrier         Neisseria meningtidis         E           Germany         1999         carrier         Neisseria meningtidis         E           Germany         1999         carrier         Neisseria meningtidis         B           227, Z1227         Demark         1974 invasive (unspecified/other) Neisseria meningtidis         B	Germany         1999         carrier         Neisseria meningitidis         W         174           Germany         1999         carrier         Neisseria meningitidis         B         19           P13; Z1213         Ghana         1973         invasive (unspecified/detr)         Neisseria meningitidis         A         4           Germany         1999         carrier         Neisseria meningitidis         NG         198           227; Z1227         Demark         1974         invasive (unspecified/det)         NGisseria meningitidis         A         5	Germany         1999         carrier         Nelsseria meninglitidis         W         174         67-174 complex           Germany         1999         carrier         Nelsseria meninglitidis         B         19         ST-18 complex           1213,21213         Gana         1973 invasive (unspecified/otter)         Nelsseria meninglitidis         A         ST-48 complex/budgroup IV           Germany         1999         carrier         Nelsseria meninglitidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meninglitidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meninglitidis         E         60         ST-60 complex           Germany         1999         carrier         Nelsseria meninglitidis         E         60         ST-60 complex           Germany         1999         carrier         Nelsseria meninglitidis         B         32         ST-32 complex/ST-32 complex           Germany         1999         carrier         Nelsseria meninglitidis         B         5         ST-50 complex/SU-32 complex           Cermany         1994         carrier         Nelsseria meninglitidis         B         5 <t< td=""><td>Germany         1999         carrier         Nelsseria meningtidis         W         174         ST-174 complex           Germany         1999         carrier         Nelsseria meningtidis         B         19         ST-18 complex           213,21213         Gramany         1973         invasive (unspecified/other)         Nelsseria meningtidis         A         4         ST-18 complex/subgroup IV         7           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         E         60         ST-60 complex           Germany         1999         carrier         Nelsseria meningtidis         B         32         ST-32 complex/EF-5 complex           Cermany         1999         carrier         Nelsseria meningtidis         B         32         ST-32 complex/EF-5 complex           227.21227         Demark         1974         invasibe (unspecified/duer)</td><td>Germany         1999         carrier         Neisseria meningtildis         W         174         ST-174 complex           Germany         1999         carrier         Neisseria meningtildis         B         19         ST-18 complex           P13, Z1213         Germany         1973         invasive (unspecified/other)         Neisseria meningtildis         A         ST-48 complex/subgroup IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         NG         198         ST-198 complex           Germany         1999         carrier         Neisseria meningtildis         NG         198         ST-198 complex           Germany         1999         carrier         Neisseria meningtildis         R         6         6         ST-60 complex           Germany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7</td></t<>	Germany         1999         carrier         Nelsseria meningtidis         W         174         ST-174 complex           Germany         1999         carrier         Nelsseria meningtidis         B         19         ST-18 complex           213,21213         Gramany         1973         invasive (unspecified/other)         Nelsseria meningtidis         A         4         ST-18 complex/subgroup IV         7           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         NG         198         ST-198 complex           Germany         1999         carrier         Nelsseria meningtidis         E         60         ST-60 complex           Germany         1999         carrier         Nelsseria meningtidis         B         32         ST-32 complex/EF-5 complex           Cermany         1999         carrier         Nelsseria meningtidis         B         32         ST-32 complex/EF-5 complex           227.21227         Demark         1974         invasibe (unspecified/duer)	Germany         1999         carrier         Neisseria meningtildis         W         174         ST-174 complex           Germany         1999         carrier         Neisseria meningtildis         B         19         ST-18 complex           P13, Z1213         Germany         1973         invasive (unspecified/other)         Neisseria meningtildis         A         ST-48 complex/subgroup IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         NG         198         ST-198 complex           Germany         1999         carrier         Neisseria meningtildis         NG         198         ST-198 complex           Germany         1999         carrier         Neisseria meningtildis         R         6         6         ST-60 complex           Germany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7         13-1           Cermany         1999         carrier         Neisseria meningtildis         B         32         ST-32 complex/SUBCPUD IV         7

Select the isolate or profile records to analyse - these will be pre-selected if you accessed the plugin following a query. Select the loci to include either directly within the loci list and/or using the schemes tree.



#### Click submit.

Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
	Toggle: i
Export allele sequences in XMFA/concatenated FASTA formats - Neisseria PubMLST	
This analysis has been submitted to the job queue.	
Please be aware that this job may take a long time depending on the number of sequences to align and how busy the server is. Alignment of hundreds of sequences can take many hours!	
Follow the progress of this job and view the output.	
Please note that the % complete value will only update after the alignment of each locus.	

The job will be submitted to the job queue. Click the link to follow the progress and download the resulting files. Sequences will be export in XMFA and FASTA file formats.

Prr17.1/ILST         Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications         Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions							
Job status viewer							
Status							
Job id:         BIGSdb_29565_1405601815_9307           Submit time         2014-07-17 13:56:55           Status:         finished           Start time:         2014-07-17 13:57:10           Progress:         100%           Stop time:         2014-07-17 13:57:23           Total time:         12 seconds							
Output   XMFA output file (not aligned)  Concatenated FASTA (not aligned)  Tar file containing output files							
Please note that job results will remain on the server for 7 days.							

### 13.2.1 Aligning sequences

By default, sequences will be exported unaligned - this is very quick since no processing is required. You can choose to align the sequences by checking the 'Align sequences' checkbox.

Options
options
If both allele designations and tagged sequences
exist for a locus, choose how you want these handled: 👔
Our contract of the sequences tagged from the bin
Use allele sequence retrieved from external database
🔽 De net instade en success with see bland flagend (defined allales will affil be word).
Do not include sequences with problem flagged (defined alleles will still be used)
Do not include incomplete sequences
Include 0 – bp flanking sequence i
Align sequences
Aligner: MAFFT 🚽
I ranslate sequences
Concatenate in frame

You can also choose to use MUSCLE or MAFFT as the aligner. MAFFT is the default choice and is usually much quicker than MUSCLE. Both produce comparable results.

# 13.3 Contig export

The contig export plugin can be accessed by clicking the 'Contigs' link in the Export section of the contents page of isolate databases.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions	
	t the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition submitted to this database and consequently it should be noted that it does not represent a population
Browse database - peruse all records.     Search by combinations of loci (profiles) - including partial matching.     List query - find isolates by matching a field to an entered list.     Projects - main projects defined in database.	Option settings • Set general options - including isolate table field handling. • Set display and query options for locus, schemes or scheme fields. • Manage submissions • Manage submissions • Manage submissions • Manage submissions • Manage submissions • Manage submissions
Breakdown     Single field     Two field     Unique combinations     Scheme and alfeles     Publications     Sequence bin     Sequence bin	Miscellaneous Codon usage Presence/absence status of loci Genome comparator BLAST

Alternatively, it can be accessed following a query by clicking the 'Contigs' button in the Export section at the bottom of the results table.

UK         2000         invasive (unspecified/other)         Neisseria meningitidis         W         1101         ST-22 complex         Complex           R81         1975         invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex/subgroup III         20         9         F3-1           R242007         UK         2000         invasive (unspecified/other)         Neisseria meningitidis         B         1102         ST-5 complex/subgroup III         20         9         F3-1           R242007         UK         2000         invasive (unspecified/other)         Neisseria meningitidis         W         114         ST-2 complex/subgroup III         20         9         F3-1           R242007         Czech Republic         1984         invasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup III         18-1         3         F5-1           190/89         Czech Republic         1989         invasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup III         5-2         10         F5-1           190/89         Czech Republic         1989         invasive (unspecified/other)         Neisseria meningitidis         X         117         <	M00240227		LIK										
7891       B54; 21054       Finland       1975       invasive (unspecified/other)       Neisseria meningitidis       A       5       ST-5 complex/subgroup III       20       9       F3-1         242007       Czech Republic       1944       invasive (unspecified/other)       Neisseria meningitidis       B       1102       ST-18 complex       14         24784       Czech Republic       1944       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-12 complex/subgroup III       18-1       3       F5-1         129       B92; 21092       Germany       1964       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup III       18-1       3       F5-6         129       B92; 21092       Czech Republic       1989       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup III       5-2       10       F5-6         120/95       Czech Republic       1999       carrier       Neisseria meningitidis       X       117       14       14       14       14       14       14       14       14       14       14       14       11       15       11       11       Germany       1999       ca							-	В			7	16	
UK         2000         invasive (unspecified/other)         Neisseria meningitidis         B         1102         ST-18 complex         14           2184         Czech Republic         1984         invasive (unspecified/other)         Neisseria meningitidis         W         114         ST-22 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex         Image: ST-18 complex							-						
22184       Czech Republic 1984 invasive (unspecified/other) Neisseria meningitidis       W       114       ST-2 complex       V         5748       B92; Z1092       Germany 1984 invasive (unspecified/other) Neisseria meningitidis       A       1       ST-1 complex/subgroup I/II       18-1       3       F5-1         199       B92; Z1092       Germany 1984 invasive (unspecified/other) Neisseria meningitidis       A       1       ST-1 complex/subgroup I/II       18-1       3       F5-1         190/89       Czech Republic 1989 invasive (unspecified/other) Neisseria meningitidis       A       1       ST-1 complex/subgroup I/II       6-2       10       F5-3         109/95       Czech Republic 1999 invasive (unspecified/other) Neisseria meningitidis       A       1       ST-1 complex/subgroup I/II       6-2       10       F5-1         11       Germany 1999 carrier       Neisseria meningitidis       X       11       ST-1 complex/subgroup I/II       6-2       10       F5-1         2       Germany 1999 carrier       Neisseria meningitidis       B       854       ST-18 complex       -       -         3131       B213; Z1213       Ghana       1973 invasive (unspecified/other) Neisseria meningitidis       A       4       ST-4 complex/subgroup I/V       7       13-1       F1-5	7891	B54; Z1054									20		F3-1
3748       B73; 21073       Canada       1971       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I       18-1       3       F5-1         129       B92; 21092       Germany       1994       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/II       5-2       10       F3-6         139M       B99; 21099       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/II       5-2       10       F5-1         20/95       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       X       117       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14								В	1102	ST-18 complex		14	
129       B92; 21092       Germany       1964       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/Subgroup /III       5-2       10       F3-6         190/89       DS9; 21099       Philippines       1986       Neisseria meningitidis       B       1015       ST-32       complex/ET-5 complex       7       16         20/95       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       X       11       ST-1 complex/Subgroup /III       5-2       10       F5-1         20/95       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       X       117       ST-1 complex/Subgroup /III       5-2       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       15       14       14       15       14       15       14       15       14       15       14       15       14       14	0021/84		Czech Republic	1984	invasive (unspecified/other)	Neisseria r	meningitidis	W	114	ST-22 complex			
99089       Czech Republic 1989 invasive (unspecified/other)       Neisseria meningitidis       B       1015       ST-32 complex/ET-5 complex       7       16         133M       B99;21099       Czech Republic 1989 invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-10 complex/ET-5 complex       7       16         120/95       Czech Republic 1989 invasive (unspecified/other)       Neisseria meningitidis       X       11       ST-10 complex/subgroup //li       6-2       10       F5-1         1       Germany       1999       carrier       Neisseria meningitidis       E       864       14         2       Germany       1999       carrier       Neisseria meningitidis       B       118       ST-18 complex       14         3       Germany       1999       carrier       Neisseria meningitidis       B       19       ST-18 complex       14         4       Germany       1999       carrier       Neisseria meningitidis       A       4       ST-4 complex/subgroup //       7       13-1       F1-5         5       Germany       1999       carrier       Neisseria meningitidis       A       4       ST-4 complex/subgroup //       7       13-1       F1-5         6       German	6748	B73; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria r	meningitidis	Α	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
B99, Z1099       Philippines       1968       Neisseria meningitidis       A       1       ST-1 complex/subgroup/I/I       5-2       10       F5-1         20/95       Czech Republic 1995       invasive (unspecified/other)       Neisseria meningitidis       X       117       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14       14	129	B92; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria r	meningitidis	Α	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
20/95       Czech Republic 1995       invasive (unspecified/other)       Neisseria meningitidis       X       117       14         1       Germany       1999       carrier       Neisseria meningitidis       E       864       14         2       Germany       1999       carrier       Neisseria meningitidis       E       864       14       14         3       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       14         4       Germany       1999       carrier       Neisseria meningitidis       W       174       ST-174 complex       14         3131       B213; Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis       A       4       574 complex/subgroup IV       7       13-1       F1-5         5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       14         6       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       14         7       Germany       1999       carrier       Neisseria meningitidis       8       32       ST-32 complex/ET-5 complex	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria r	meningitidis	В	1015	ST-32 complex/ET-5 complex	7	16	
1       Germany       1999       carrier       Neisseria meningitidis       E       864       Image: State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State Sta	139M	B99; Z1099	Philippines	1968		Neisseria r	meningitidis	Α	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
2       Germany       1999       carrier       Neisseria meningitidis       B       854       ST-18 complex       Image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the ima	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria r	meningitidis	Х	117			14	
3       Germany       1999       carrier       Neisseria meningitidis       W       174       ST-174 complex       Image: State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State State Sta	1		Germany	1999	carrier	Neisseria r	meningitidis	E	864				
4       Germany       1999       carrier       Neisseria meningitidis       B       19       ST-18 complex         3131       B213; Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis       A       4       ST-4 complex/subgroup IV       7       13-1       F1-5         5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-18 complex       7         6       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       7         7       Germany       1999       carrier       Neisseria meningitidis       B       32       ST-30 complex       7         8       Germany       1999       carrier       Neisseria meningitidis       B       32       ST-30 complex         4355       B227, Z1227       Denmark       1974       invasive (unspecified/other)       Neisseria meningitidis       B       9       ST-334 complex       7         9       Germany       1999       carrier       Neisseria meningitidis       B       9       37.334 complex       9         9       Germany       1999       carrier       Neisseria meningitidis       B <td>2</td> <td></td> <td>Germany</td> <td>1999</td> <td>carrier</td> <td>Neisseria r</td> <td>meningitidis</td> <td>В</td> <td>854</td> <td>ST-18 complex</td> <td></td> <td></td> <td></td>	2		Germany	1999	carrier	Neisseria r	meningitidis	В	854	ST-18 complex			
3131       B213; Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis       A       4       ST-4 complex/subgroup IV       7       13-1       F1-5         5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       Image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Comp	3		Germany	1999	carrier	Neisseria r	meningitidis	W	174	ST-174 complex			
5       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       Image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the image: Complex in the i	4		Germany	1999	carrier	Neisseria r	meningitidis	в	19	ST-18 complex			
6       Germany       1999       carrier       Neisseria meningitidis       NG       198       ST-198 complex       198       ST-198 complex         7       Germany       1999       carrier       Neisseria meningitidis       E       60       ST-60 complex       198       ST-128 complex       198       ST-128 complex       198       ST-128 complex       198       ST-128 complex       198       ST-328 complex       199       199       carrier       Neisseria meningitidis       A       5       ST-52 complex/Stobusproup III       5-1       9       F3-1         9       Germany       1999       carrier       Neisseria meningitidis       B       930       ST-334 complex       F3-1       9       F3-1         is tools:       Stobuspecified/other       Neisseria meningitidis       B       930       ST-334 complex       Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobuspecified/Stobu	S3131	B213; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria r	meningitidis	Α	4	ST-4 complex/subgroup IV	7	13-1	F1-5
7     Germany     1999     carrier     Neisseria meningitidis     E     60     ST-60 complex       8     Germany     1999     carrier     Neisseria meningitidis     B     32     ST-32 complex/ET-5 complex       4355     B227; Z1227     Denmark     1974     invasive (unspecified/other)     Neisseria meningitidis     A     5     ST-52 complex/Subgroup III     5-1     9     F3-1       9     Germany     1999     carrier     Neisseria meningitidis     B     930     ST-334 complex     F1-9     F3-1	5		Germany	1999	carrier	Neisseria r	meningitidis	NG	198	ST-198 complex			
8     Germany     1999     carrier     Neisseria meningitidis     B     32     ST-32 complex/ET-5 complex       4255     B227; Z1227     Denmark     1974     invasive (unspecified/other)     Neisseria meningitidis     A     5     ST-5 complex/subgroup III     5-1     9     F3-1       9     Germany     1999     carrier     Neisseria meningitidis     B     930     ST-334 complex	6		Germany	1999	carrier	Neisseria r	meningitidis	NG	198	ST-198 complex			
4355       B227; Z1227       Denmark       1974       Invasive (unspecified/other)       Neisseria meningitidis       A       5       ST-5 complex/subgroup III       5-1       9       F3-1         9       Carrier       Neisseria meningitidis       B       930       ST-334 complex       F3-1       9       F3-1         is tools:       Image: Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       State Carrier       <	7		Germany	1999	carrier	Neisseria r	meningitidis	E	60	ST-60 complex			
9 Germany 1999 carrier Neisseria meningitidis B 930 ST-334 complex	8		Germany	1999	carrier	Neisseria r	meningitidis	в	32	ST-32 complex/ET-5 complex			
is tools:	S4355	B227; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria r	meningitidis	Α	5	ST-5 complex/subgroup III	5-1	9	F3-1
	9		Germany	1999	carrier	Neisseria r	meningitidis	В	930	ST-334 complex			
WII. Fields Workeid Polymorphicisties Combinations Schemes/alleles Publications Sequence bin Tag status	S4355 9 lysis tools	5:	Denmark Germany	1974 1999	invasive (unspecified/other) carrier	Neisseria r Neisseria r	meningitidis meningitidis	A B	5 930	ST-5 complex/subgroup III ST-334 complex	5-1	9	F3-1
	Export: Da	ataset Conti	gs Sequences										
Dort Dataset Contigs Sequences				_									
		M00242007 0021/84 6748 129 0090/89 139M 0120/95 1 2 3 4 4 S3131 5 6 7 8 S4355 9 Iysis tools akdown: Fi nalysis: B	M00242007 002184 6748 B73; Z1073 129 B92; Z1092 0090/89 139M B99; Z1099 0120/95 1 4 S3131 B213; Z1213 5 6 7 8 S4355 B227; Z1227 9 Iysis tools: akdown: Fields Two Fi nalysis: BURST Code	M00242007         UK           002184         Czech Republic           6748         B73; Z1073         Canada           129         B92; Z1092         Germany           0090/89         Czech Republic         Czech Republic           139M         B99; Z1099         Philippines           0120/95         Czech Republic         Czech Republic           1         Germany         Czech Republic           3         Germany         Germany           4         Germany         Germany           5         Germany         Germany           6         Germany         Germany           7         Germany         Germany           8         B27; Z1227         Denmark           9         B27; Z1227         Denmark           Iysis tools:         Issee State         State           Itysis tools:         Two Field         Polymorph           BUST         Codors         Presence/A	M00242007         UK         2000           002184         Czech Republic 1984         737           6748         B73; 21073         Canada         1971           129         B92; 21092         Germany         1984           0090/89         Czech Republic 1989         1989         1989           139M         B99; 2109         Philippines         1988           0120/95         Czech Republic 1995         1         Germany         1999           2         Germany         1999         Germany         1999           3         Germany         1999         Germany         1999           4         Germany         1999         Germany         1999           5         Germany         1999         Germany         1999           6         Germany         1999         Germany         1999           7         Germany         1999         Germany         1999           84355         B227; Z1227         Denmark         1974           9         Stools:         Ivos Field         Polymorphic sites           Iysis tools:         BURST         Codors         Presence/Abernee	M00242007         UK         2000         invasive (unspecified/other)           002184         Czech Republic         1984         invasive (unspecified/other)           129         B92, Z1092         Germany         1964         invasive (unspecified/other)           129         B92, Z1092         Germany         1964         invasive (unspecified/other)           139M         B92, Z1092         Czech Republic         1985         invasive (unspecified/other)           139M         B92, Z1099         Philippines         1986         invasive (unspecified/other)           139M         B92, Z1099         Czech Republic         1995         invasive (unspecified/other)           1         Germany         1999         carrier         Germany         1999         carrier           2         Germany         1999         carrier         Germany         1999         carrier           3         Germany         1999         carrier         Germany         1999         carrier           5         Germany         1999         carrier         Germany         1999         carrier           7         Germany         1999         carrier         Germany         1999         carrier           84355	M00242007         UK         2000         invasive (unspecified/other)         Neisseria i           002184         Czech Republic         1984         invasive (unspecified/other)         Neisseria i           129         B92; Z1092         Germany         1964         invasive (unspecified/other)         Neisseria i           0090/89         Czech Republic         1989         invasive (unspecified/other)         Neisseria i           0120/95         Czech Republic         1989         invasive (unspecified/other)         Neisseria i           0120/95         Czech Republic         1995         invasive (unspecified/other)         Neisseria i           1         Germany         1999         carrier         Neisseria i           2         Germany         1999         carrier         Neisseria i           3         Germany         1999         carrier         Neisseria i           4         Germany         1999         carrier         Neisseria i           5         Germany         1999         carrier         Neisseria i           6         Germany         1999         carrier         Neisseria i           7         Germany         1999         carrier         Neisseria i           8	M00242007         UK         2000         invasive (unspecified/other)         Neisseria meningitidis           002184         Czech Republic         1984         invasive (unspecified/other)         Neisseria meningitidis           6748         B73; 21073         Canada         1984         invasive (unspecified/other)         Neisseria meningitidis           129         B92; Z1092         Germany         1964         invasive (unspecified/other)         Neisseria meningitidis           0090/89         Czech Republic         1985         invasive (unspecified/other)         Neisseria meningitidis           012005         Czech Republic         1995         invasive (unspecified/other)         Neisseria meningitidis           1         Germany         1999         carrier         Neisseria meningitidis           2         Germany         1999         carrier         Neisseria meningitidis           3         Germany         1999         carrier         Neisseria meningitidis           5         Germany         1999         carrier         Neisseria meningitidis           6         Germany         1999         carrier         Neisseria meningitidis           6         Germany         1999         carrier         Neisseria meningitidis           <	M00242007         UK         2000         invasive (unspecified/other)         Neisseria meningitidis         B           002184         Czech Republic         1984         invasive (unspecified/other)         Neisseria meningitidis         M           129         B92; Z1092         Germany         1964         invasive (unspecified/other)         Neisseria meningitidis         A           129         B92; Z1092         Germany         1964         invasive (unspecified/other)         Neisseria meningitidis         A           0090/89         DS2, Z1092         Germany         1969         invasive (unspecified/other)         Neisseria meningitidis         A           0120/95         Czech Republic         1995         invasive (unspecified/other)         Neisseria meningitidis         X           1         Germany         1999         carrier         Neisseria meningitidis         B           2         Germany         1999         carrier         Neisseria meningitidis         B           3         Germany         1999         carrier         Neisseria meningitidis         B           3         Germany         1999         carrier         Neisseria meningitidis         B           5         Germany         1999         carrier	M00242007         UK         2000         invasive (unspecified/other)         Neisseria meningitidis         B         1102           002184         Czech Republic         1984         invasive (unspecified/other)         Neisseria meningitidis         W         114           6748         B73; 21073         Canada         1971         invasive (unspecified/other)         Neisseria meningitidis         A         1           129         B92; Z1092         Germany         1964         invasive (unspecified/other)         Neisseria meningitidis         A         1           090/089         Dispecified/other)         Neisseria meningitidis         A         1         1         1020/05         2         Cerch Republic         1996         invasive (unspecified/other)         Neisseria meningitidis         X         117           1         Germany         1999         carrier         Neisseria meningitidis         B         195           3         Germany         1999         carrier         Neisseria meningitidis         B         195           4         Germany         1999         carrier         Neisseria meningitidis         M         174           5         Germany         1999         carrier         Neisseria meningitidis         M <td>M00242007       UK       2000       invasive (unspecified/other)       Neisseria meningitidis       B       1102       ST-18 complex         002184       Czech Republic       1984       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-12 complex/subgroup I/I         129       B92, Z1092       Germany       1964       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I         0090/89       Czech Republic       1996       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I         0120045       Czech Republic       1996       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I         0120045       Czech Republic       1996       carrier       Neisseria meningitidis       X       117         1       Germany       1999       carrier       Neisseria meningitidis       B       15       ST-18 complex         3       Germany       1999       carrier       Neisseria meningitidis       B       17       ST-14       complex         3131       B213, Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis</td> <td>M00242007       UK       2000       invasive (unspecified/other)       Neisseria meningitidis       B       1102       ST-18 complex         002184       Czech Republic       1984       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-12 complex/subgroup I/I       18-1         129       B92, Z1092       Germany       1964       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I       5-2         0090/89       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I       5-2         012005       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       X       117         1       Germany       1996       carrier       Neisseria meningitidis       X       117         1       Germany       1999       carrier       Neisseria meningitidis       B       195       ST-18 complex       7         3       Germany       1999       carrier       Neisseria meningitidis       B       19       ST-14       complex       1       15       ST-18 complex       1       1       ST-14 complex/subgroup I/I       7       1</td> <td>M00242007         UK         2000         Imasive (unspecified/other)         Neisseria meningitidis         B         1102         ST-18 complex         14           0021/84         Czech Republic         1984         imasive (unspecified/other)         Neisseria meningitidis         W         114         ST-22 complex         14           0021/84         Czech Republic         1984         imasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup //ll         18-1         3           129         B92,21092         Germany         1964         imasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup //ll         5-2         10           0190/89         Czech Republic         1986         Neisseria meningitidis         A         1         ST-1 complex/subgroup //ll         5-2         10           0120/95         Czech Republic         1995         invasive (unspecified/other)         Neisseria meningitidis         X         117         14           1         Germany         1999         carrier         Neisseria meningitidis         B         184         ST-18 complex         14           2         Germany         1999         carrier         Neisseria meningiti</td>	M00242007       UK       2000       invasive (unspecified/other)       Neisseria meningitidis       B       1102       ST-18 complex         002184       Czech Republic       1984       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-12 complex/subgroup I/I         129       B92, Z1092       Germany       1964       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I         0090/89       Czech Republic       1996       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I         0120045       Czech Republic       1996       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I         0120045       Czech Republic       1996       carrier       Neisseria meningitidis       X       117         1       Germany       1999       carrier       Neisseria meningitidis       B       15       ST-18 complex         3       Germany       1999       carrier       Neisseria meningitidis       B       17       ST-14       complex         3131       B213, Z1213       Ghana       1973       invasive (unspecified/other)       Neisseria meningitidis	M00242007       UK       2000       invasive (unspecified/other)       Neisseria meningitidis       B       1102       ST-18 complex         002184       Czech Republic       1984       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-12 complex/subgroup I/I       18-1         129       B92, Z1092       Germany       1964       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I       5-2         0090/89       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       A       1       ST-1 complex/subgroup I/I       5-2         012005       Czech Republic       1995       invasive (unspecified/other)       Neisseria meningitidis       X       117         1       Germany       1996       carrier       Neisseria meningitidis       X       117         1       Germany       1999       carrier       Neisseria meningitidis       B       195       ST-18 complex       7         3       Germany       1999       carrier       Neisseria meningitidis       B       19       ST-14       complex       1       15       ST-18 complex       1       1       ST-14 complex/subgroup I/I       7       1	M00242007         UK         2000         Imasive (unspecified/other)         Neisseria meningitidis         B         1102         ST-18 complex         14           0021/84         Czech Republic         1984         imasive (unspecified/other)         Neisseria meningitidis         W         114         ST-22 complex         14           0021/84         Czech Republic         1984         imasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup //ll         18-1         3           129         B92,21092         Germany         1964         imasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex/subgroup //ll         5-2         10           0190/89         Czech Republic         1986         Neisseria meningitidis         A         1         ST-1 complex/subgroup //ll         5-2         10           0120/95         Czech Republic         1995         invasive (unspecified/other)         Neisseria meningitidis         X         117         14           1         Germany         1999         carrier         Neisseria meningitidis         B         184         ST-18 complex         14           2         Germany         1999         carrier         Neisseria meningiti

Select the isolates for which you wish to export contig data for. If the export function was accessed following a query, isolates returned in the query will be pre-selected.

	Profile/ST   List Is   Scheme/alleles   Publications   Options   Profiles/sequences definitions			
			Help 🖉	Toggle: 🚯
Contig analysis and expo	ort			
calculated by adding up the length of all loci t the length of the contig.	which contigs are associated - use Ctrl or Shift to make multiple lagged within the contig - if these loci overlap then the total tagge	ed length will be reported		
Isolates 1) A4/M1027 2) 120M 7) 7891 10) 6748 11) 129E 13) 139M 19) S3131 24) S4355 All None	Options Identify contigs with >= 0 • % of sequence untagged FASTA header line: original designation • 0 Action Reset Submit	Filter by Sequence method: Project Experiment Minimum length:	• 0 • 0 • 0	• 0

At its simplest, press submit.

A table will be produced with download links. Clicking these will produce the contigs in FASTA format.

- 15	solates ——				- Options	- Filter by		
1) A	A4/M1027				Identify contigs with >= 0	Sequence method:	- 0	
	120M				FASTA header line: original designation - 1	Project:		-
	7891 ) 6748				······································	Experiment:	• 0	
	129E					Minimum length:		
	139M					Minimum lengul.	<b>-</b> 0	
	S3131				-Action			
				<b>T</b>				
:4)	S4355				Reset Submit			
(4)	S4355	JI None	)		Reset Submit			
24)		II None	)		Reset Submit			
:4)		I None	)		Reset Submit			
(4)		JI None	)		Reset Submit			
			- -	ength u				
	A		juence le	-	untagged			
nt	A	=0% seq	uence le matching	contigs	untagged			
nt	tigs with ≻ isolate ✦ co	=0% seq ontigs <del>\$</del>	uence le matching count ¢ d	contigs Iownloa	untagged s non-matching contigs d count \$ download			
ont	tigs with > isolate ✦ co A4/M1027	=0% seq ontigs \$ 364	uence le matching count + d 364	contigs lownloa	untagged s non-matching contigs d count ¢ download 0			
nt	tigs with ≻ isolate ✦ co	=0% seq ontigs <del>\$</del>	uence le matching count ¢ d	contigs Iownloa	untagged s non-matching contigs d count \$ download			

You can also download all the data in a tar file by clicking the 'Batch download' link.

	olates ——				Options				
	4/M1027			▲ Id	entify contigs with >= 0	<ul> <li>% of sequence untagged</li> </ul>	Sequence method:	- (	•
	20M /891			E F/	ASTA header line: original	designation 👻 🕕	Project:		
	6748						Experiment:	<b>v</b> 0	
	129E						Minimum length:	• 0	
	139M S3131				Action				
	S4355			-	Action				
		All None	٦ ۲		Reset Submit				
nt	igs with >	•=0% sec	luence le	ength ui	ntagged				
	-			-	ntagged				
	igs with > isolate ≑ c			contigs	non-matching contigs				
÷	-		matching	contigs	non-matching contigs				
÷	isolate 🗢 c	contigs 🗢	matching count <del>\$</del> d	contigs lownload	non-matching contigs count \$ download				
÷	isolate	contigs \$ 364	matching count ¢ d 364	contigs lownload	non-matching contigs count 💠 download 0				

#### 13.3.1 Filtering by tagged status of contigs

You can also export contigs based on the percentage of the sequence that has been tagged. This is useful to find sequences to target for gene discovery.

In order to export contigs where at least half the sequence has been tagged (and also the remaining contigs in a separate file), select '50' in the dropdown box for %untagged.

Options			
Identify contigs with	>= 50	🚽 % of seq	uence untagged
FASTA header line:	original	designation	<b>▼</b> [i

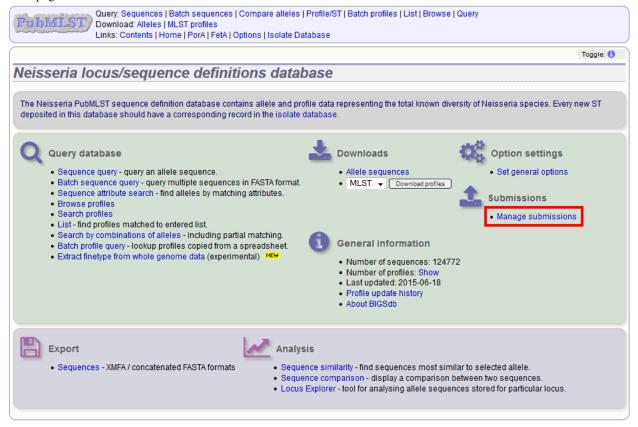
The resulting table has two download links for each isolate, one for contigs matching the condition, and one for contigs that don't match.

Con	Contigs with >=50% sequence length untagged									
id 🔺	icolato 🔺	contigo 🔺	matchin	g contigs	non-matching contig					
id 🗢 isolate 🗢		conugs <del>-</del>	count 🔶	download	count 🔶	download				
1	A4/M1027	364	163	*	201	*				
2	120M	359	81	*	278	*				
7	7891	199	48	*	151	*				
10	6748	652	393	<u>*</u>	259	*				

# Submitting data using the submission system

The automated submission system allows users to submit data (new alleles, profiles, or isolates) to the database curators for assignment and upload to the database. The submission system is enabled on a per-database basis so will not always be available.

If the system is enabled, new submissions can be made by clicking the 'Manage submissions' link on the database front page.



### 14.1 Registering a user account

You must have an account for the appropriate database in order to use the submission system. This will need to be set up by a curator, so contact them in the first instance.

# 14.2 Allele submission

New allele data can only be submitted from within the appropriate sequence definition database. Submissions consist of one or more new allele sequences for a single locus. You will need to create separate submissions for each locus - this is because different loci may be handled by different curators.

### 14.2.1 Start

Click the 'alleles' link under submission type on the submission management page.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Logged in: Keith Jolley (keith). House out   Change password	
Manage submissions	
Submit new data	
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.	
Submission type:	
alleles     MLST profiles	
Return to index page	

### 14.2.2 Select the submission locus

Select the locus from the locus list box:

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   E Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	Batch profiles   List   Browse   Query
Logged in: Keith Jolley (keith2). @Log out   Change password	Help 🖉
Manage submissions	
Submit new alleles You need to make a separate submission for each locus for which you have new alleles - th any number of new sequences for a single locus as one submission. Sequences should be Filter loci by scheme Filter Capsule Genetic Information Processing Hetabolism Loci not in schemes Loci not in schemes Filter (NEIS1279) acn8 (NEIS1279) acn8 (NEIS1279) acn8 (NEIS1279) acn8 (NEIS1279) acn8 (NEIS129) acn8	

The locus list may be very long in some databases. It may be possible to filter these to those belonging to specific schemes. If the scheme tree is shown, select the appropriate scheme, e.g. 'MLST' and click 'Filter'.

Query: Sequences   Batch sequences Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA	 atch profiles   List   Browse   Query	
Logged in: Keith Jolley (keith). DLog out   Change password		
Manage submissions		
Submit new alleles You need to make a separate submission for each locus number of new sequences for a single locus as one sub Filter loci by scheme All loci Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Capsule Ca		

The locus list is now constrained making selection easier.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database		
Logged in: Keith Jolley (keith). GLog out   Change password		
Manage submissions		
		alleles - this is because different loci may have different curators. You can submit any nuld be trimmed to the correct start/end sites for the selected locus. Sequence details technology! Illumina read length:! 100-199 coverage:! 20-49x assembly! de novo assembly! de novo FASTA or single sequence Action Submit

#### 14.2.3 Enter details of sequencing method

There are a number of fields that must be filled in so that the curator knows how the sequence was obtained:

- technology the sequncing platform used, allowed values are:
  - 454
  - Illumina
  - Ion Torrent
  - PacBio
  - Oxford Nanopore
  - Sanger
  - Solexa
  - SOLiD
  - other
  - unknown
- read length this is the length of sequencing reads. This is a required field for Illumina data, and not relevant to Sanger sequencing. Allowed values are:
  - <100
  - 100-199
  - 200-299
  - 300-499
  - >500

- coverage the mean number of reads covering each nucleotide position of the sequence. This is not relevant to Sanger sequencing, Allowed values are:
  - <20x
  - 20-49x
  - 50-99x
  - >100x
- assembly the means of generating the submitted sequence from the sequencing reads. Allowed values are:
  - de novo
  - mapped
- assembly software this is a free text field where you should enter the name of the software used to generate the submitted sequence.

#### 14.2.4 Paste in sequence(s)

Paste in the new variant sequences to the box. This can either be a stand- alone sequence or multiple sequences in FASTA format. The sequences must be trimmed to the start and end points of the loci - check existing allele definitions if in doubt. The submission is likely to be rejected if sequences are not trimmed. Click submit.

Query:         Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download:         Alleles   MLST profiles           Links:         Contents   Home   PorA   FetA   Options   Isolate Database		
Logged in: Keith Jolley (keith). I Log out   Change password		
Manage submissions		
		alleles - this is because different loci may have different curators. You can submit any ould be trimmed to the correct start/end sites for the selected locus.  Sequence details  technology! Illumina  read length! 100-199  coverage! 20-49x assembly! de novo assembly! de novo assembly software! Velvet  FASTA or single sequence Action Submit  Submit

The system will perform some basic checks on the submitted sequences. If any of the sequences have been defined previously they must be removed from the submission before you can proceed. Curators do not want to waste their time dealing with previously defined sequences.

PTTIMIST Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database			
Logged in: Keith Jolley (keith). HoLog out   Change password			
Manage submissions			
Error: Sequence 'NM432' has already been defined as abcZ-3. Submit new alleles You need to make a separate submission for each locus for which you new sequences for a single locus as one submission. Sequences sh — Filter loci by scheme — Section 2015	hould be trimmed to the correct start/end sites for the selected locu elect locus - Sequence details technology:! Illumina • read length:! 100-199 • nC coverage:! 20-49x • nC assembly:! de novo •		

Assuming the preliminary checks have passed you will then be able to add additional information to your submission.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   Lis Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	st   Browse   Query
Logged in: Keith Jolley (keith2). @Log out   Change password	Help 🖉
Manage submissions	
Submission: BIGSdb_20150709114223_7557_43592         Sequences       Sequences         You are submitting the following abcZ sequences: Download       Image: Colspan="2">Sequence         Identifier       Length       Sequence         NM322       433       TITGATACTGTTECCGAACG       GCGGATTGTCGAACTCGATC pending         NM21       433       TITGATACCGTTECCGAAGG       GCGGATTGTCGAACTCGATC pending	Sequence details technology: Illumina  read length: 100-199  coverage: 20-49x  assembly: de novo  assembly: Velvet
E-mail     Opdates will be sent to kajolley@gmail.com.     Finalize submission!     Finalize submississe     Finalize submission!     Finaliz	idd an explanatory note so that they can be linked to
Browse No files selected. Upload files Messages	

### 14.2.5 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Append'.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   L           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database	ist   Browse   Query
Logged in: Keith Jolley (keith2). (hLog out   Change password	Help 🗹
Manage submissions	
Submission: BIG Sdb_20150709114223_7557_43592 Sequences You are submitting the following abcZ sequences: Download MM322 433 TTTGATACTGTTGCCGAACG GCGGATTGTCGAACTCGATC pending NM21 433 TTTGATACCGTTGCCGAACG GCGGATTGTCGAACTCGATC pending E-mail 433 TTTGATACCGTTGCCGAACG GCGGATTGTCGAACTCGATC pending E-mail submission updates Supporting files Please upload any supporting files required for curation. Ensure that these are named unambiguously or the appropriate submission item. Individual filesize is limited to 32 MB. Browse. No files selected. Upload files Messages The sequence variant NM322 has been seen in 3 isolates. 	Sequence details technology! Illumina read lengthi! 100-199 coverage: 20-49x assembly: de novo assembly software: Velvet

The message will be attached. A curator may respond to the message and attach their own, with the full conversation becoming part of the submission record.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database		
Logged in: Keith Jolley (keith2). In Log out   Change password	Help 🗗	
Manage submissions		
Submission: BIGSdb_20150709114223_7557_43592         Sequences       Sequences: Download         You are submitting the following abcZ sequences: Download       Email         Identifier Length       Sequence       Status         MM322       433       TITGATACTGTTGCCGAACG       GCGGATTGTCGAACTCGATC pending         NM21       433       TITGATACCGTTGCCGAAGG       GCGGATTGTCGAACTCGATC pending         E-mail       Action	Sequence details technology! Illumina • read length:! 100-199 • coverage:! 20-49x • assembly! de novo • assembly software:! Velvet	
<ul> <li>E-mail submission updates</li> <li>Supporting files</li> <li>Please upload any supporting files required for curation. Ensure that these are named unambiguously or a appropriate submission item. Individual filesize is limited to 32 MB.</li> </ul>	dd an explanatory note so that they can be linked to the	
Browse No files selected. Upload files		
Messages         Timestamp       User       Message         2015-07-09 10:44:46+00       Keith Jolley       The sequence variant NM322 has been seen in 3 isolates.        :         Message:         Append		

### 14.2.6 Add supporting files

Some submissions will require the attachment of supporting files. This will depend on the policies of the individual databases. Sequences determined by Sanger sequencing should normally have forward and reverse trace files attached.

Files can be added to the submission by clicking the 'Browse' button in the 'Supporting files' section.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database		
Logged in: Keith Jolley (keith2). IPLog out   Change password	Help 🗗	
Manage submissions		
Submission: BIGSdb_20150709114223_7557_43592         Sequences         You are submitting the following abcZ sequences: Download Image:	Sequence details technology! Illumina • read length:! 100-199 • coverage:! 20-49x • assembly!! de novo • assembly software:! Velvet	
Supporting files         Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.         Browse No files selected.       Upload files         Messages       Timestamp       User         2015-07-09 10:44:46+00       Keith Jolley       The sequence variant NM322 has been seen in 3 isolates.		

Select the file in the selection box, then click 'Upload files'.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database		
Logged in: Keith Jolley (keith2). ( Log out   Change password	Help 🗹	
Manage submissions		
Submission: BIG Sdb_20150709114223_7557_43592         Sequences         You are submitting the following abcZ sequences: Download         Identifier Length       Sequence       Status       Assigned allele         NM322       433       TITGATACTGTTGCCGAACG       GCGGATTGTCGAACTCGATC       pending         NM21       433       TITGATACCGTTGCCGAAGG       GCGGATTGTCGAACTCGATC       pending         E-mail       Action	Sequence details technology:1 Illumina • read length:1 100-199 • coverage:1 20-49x • assembly:1 de novo • assembly software:1 Velvet	
✓ E-mail submission updates         Supporting files         Please upload any supporting files required for curation. Ensure that these are named unambiguously or ad appropriate submission item. Individual filesize is limited to 32 MB.         Browse abcZ.ace       Upload files         Messages	d an explanatory note so that they can be linked to the	

The file will be uploaded and shown in a table.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   Lis Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	t   Browse   Query		
Logged in: Keith Jolley (keith2). (+Log out   Change password			Help 🖉
Manage submissions			
Submission: BIG\$db_20150709114223_7557_43592			
Sequences	— Sequence details —		
You are submitting the following abcZ sequences: Download	technology:	Illumina	•
	read length:!	100-199 👻	
Identifier         Length         Sequence         Status         Assigned allele           NM322         433         TTTGATACTGTTGCCGAACG         CCGGATTGTCGAACTCGATC         pending		20-49x 👻	
NM21 433 TTTGATACCGTTGCCGAAGG GCGGATTGTCGAACTCGATC pending	assembly: assembly software:	de novo 👻	
	assembly sollware.	veivel	
E-mail     Department of kajolley@gmail.com.     Einalize submission			
E-mail submission updates			
Supporting files			
Please upload any supporting files required for curation. Ensure that these are named unambiguously or a appropriate submission item. Individual filesize is limited to 32 MB.	id an explanatory note s	so that they can be	linked to the
Browse No files selected. Upload files			
Filename Size Delete			
Delete selected files			
Messages			
Timestamp User Message 2015-07-09 10:44:46+00 Keith Jollev The sequence variant NM322 has been seen in 3 isolates.			
Message: Append			

Files can be removed from the submission by checking the appropriate 'Delete' box and clicking 'Delete selected files'.

#### 14.2.7 Finalize submission

Make sure the 'E-mail submission updates' box is checked if you wish to receive E-mail notification of the result of your submission. This setting is remembered between submissions.

Click 'Finalize submission!'.

PrrDMLST Ownload: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	st   Browse   Query
Logged in: Keith Jolley (keith2). @Log out   Change password	Help 🖓
Manage submissions	
Submission: BIGSdb_20150709114223_7557_43592	- Sequence details - E-mail
You are submitting the following abcZ sequences: Download	technology: Illumina Updates will be sent to kajolley@gmail.com.
Identifier         Length         Sequence         Status         Assigned allele           NM322         433         TITIGATACTGTIGCCGAACG         GCGGATIGTCGAACTCGATC         pending           NM21         433         TITIGATACCGTIGCCGAAGG         GCGGATIGTCGAACTCGATC         pending	read length: 100-199 ↓ coverage: 20-49x ↓ assembly: de novo ↓ assembly: de novo ↓ Finalize submission!
Supporting files	
Please upload any supporting files required for curation. Ensure that these are named unambiguously or 32 MB.	add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to
Browse No files selected. Upload files	
Filename Size Delete abcZ.ace 1.7 MB	
Delete selected files	
Messages Timestamp User Message	
2015-07-09 10:44:46+00 Keith Jolley The sequence variant NM322 has been seen in 3 isolates.	
Message: Append	

Your submission will then be listed under 'Pending submissions' on your submission page.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Logged in: Keith Jolley (keith2). @Log out   Change password	Help 🖉
Manage submissions	
Submit new data	
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.	
Submission type:	
alleles     MLST profiles	
Pending submissions	
You have submitted the following submissions that are pending curation:	
Submission id         Submitted         Updated         Type         Details           BIGSdb_20150709114223_7557_43592         2015-07-09         2015-07-09         alleles         2 abcZ sequences	
Return to index page	

# 14.3 Profile submission

#### 14.3.1 Start

**Note:** Most MLST databases on PubMLST.org require you to submit an isolate record for each new ST that you wish to be defined. In these cases, you should add the isolate name to the id field of your profile submission and make a corresponding *isolate submission* containing the allelic profile.

Click the appropriate profiles link under submission type on the submission management page.



Download the Excel submission template.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query:           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database	'
Logged in: Keith Jolley (keith). @Log out   Change password	
Manage submissions	
Submit new MLST profiles	
Paste in your profiles for assignment using the template available below.	
Download tab-delimited header for your spreadsheet - use 'Paste Special SText' to paste the data.     Download submission template (xlsx format)	
Please paste in tab-delimited text (include a field header line)	Action
	Submit

# 14.3.2 Paste in profile(s)

Fill in the template. The first column 'id' can be used to enter an identifier that is meaningful to you - it is used to report back the results but is not uploaded to the database. It can be left blank, or the entire column can be removed - in which case individual profiles will be identified by row number.

Copy and paste the entire contents of the submission worksheet. Click submit.

PubMLST	Query: Seque Download: Al Links: Conter	leles   MLST	profiles			n profiles   List   Bi	rowse   Query		
Logged in: Keith Jolley	(keith). 🗭Log ou	ut   Change pa	ssword						
Manage sub	missio	ns							
Submit new MLS Paste in your profiles Download ta Download su Please paste in 1d abcZ 8 5 3	s for assignm b-delimited he ibmission ten	eader for you nplate (xlsx fo	r spreadshe ormat)	eet - use 'Pa	C Text to paste D m 32 2 6	e the data.		- Action Submit	

Some basic checks will be performed. These include whether the profile has already been assigned and whether each allele identifier exists. The submission cannot proceed if the checks fail.

Pu	ŢŊŢĊŢ	$_{\rm ST}$ Do	wnload: All	nces   Batch eles   MLST ts   Home   F	profiles			file/ST   Batch profiles   Lis ase	st   Browse   Quer	у	
Logg	ed in: Kei	th Jolley (kei	th). 🕩Log out	Change pas	sword						
Ма	nage	subn	nissior	າຣ							
Eri	ror:										
Ro	w 3: Profi	le has alre:	ady been de	fined as ST-	10018.						
		w MLST	·								
Pas			-	ent using the							
				ader for your plate (xlsx fo		et-use 'Pa	ste Special (	Text to paste the data.			
	Please	paste in tab	-delimited f	text (include	a field head	der line) —					
i	.d	abcZ 8	adk 5	aroE 32	fumC 2	gdh 6	pdhC 43	pgm 32		Submit	
		5	7	3	6	33	12	2			
		3	2	4	3	8	4	6			

Provided the checks pass, you will then be able to add additional information to your submission

# 14.3.3 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Append'.

Query: Sequences   Batch sequences   Compare alleles Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Di	
Logged in: Keith Jolley (keith2). @Log out   Change password	Help 🖒
Manage submissions	
Submission: BIGSdb_20150709115437_16625_11378         Profiles         You are submitting the following MLST profiles: Download Image: Down	E-mail Action Updates will be sent to kajolley@gmail.com. E-mail submission updates
Supporting files Please upload any supporting files required for curation. Ensure that these a submission item. Individual filesize is limited to 32 MB. Browse No files selected. Upload files Messages Corresponding isolate data has already been submitted (ids 43721-43722). 	are named unambiguously or add an explanatory note so that they can be linked to the appropriate

The message will be attached. A curator may respond to the message and attach their own, with the full conversation becoming part of the submission record.

Query: Sequences   Batch sequences   Compare alleles Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Da		
Logged in: Keith Jolley (keith2). In Log out   Change password		Help 🖉
Manage submissions		
Submission: BIGSdb_20150709115437_16625_11378  Profiles You are submitting the following MLST profiles: Download  (a) (dentifier adk abc2 aroE fumC gdh pdhC pgm Status Assigned ST Row 1 5 8 32 2 6 43 32 pending Row 2 7 5 3 6 33 12 2 pending	E-mail Updates will be sent to kajolley@gmail.com. ☑ E-mail submission updates	Action Finalize submission!
Supporting files Please upload any supporting files required for curation. Ensure that these a submission item. Individual filesize is limited to 32 MB. Browse. No files selected. Upload files	are named unambiguously or add an explanatory i	note so that they can be linked to the appropriate
Messages Timestamp User Me: 2015-07-09 10:56:00+00 Keith Jolley Corresponding isolate data has alree	ssage ady been submitted (ids 43721-43722).  Message: Append	

# 14.3.4 Add supporting files

Some submissions may require the attachment of supporting files. These files can be added to the submission by clicking the 'Browse' button in the 'Supporting files' section.

Select the file in the selection box, then click 'Upload files'.

# 14.3.5 Finalize submission

Make sure the 'E-mail submission updates' box is checked if you wish to receive E-mail notification of the result of your submission. This setting is remembered between sessions.

Query: Sequences   Batch sequences   Compare alle Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate		ny
Logged in: Keith Jolley (keith2). @Log out   Change password		Help 🖓
Manage submissions		
Submission: BIGSdb_20150709115437_16625_11378		
Profiles	E-mail	Action
You are submitting the following MLST profiles: Download Identifier adk abc2 aroE fumC gdh pdhC pgm Status Assigned S Row 1 5 8 32 2 6 43 32 pending Row 2 7 5 3 6 33 12 2 pending Supporting files Please upload any supporting files required for curation. Ensure that these submission item. Individual filesize is limited to 32 MB.		Finalize submission!
Browse No files selected. Upload files Messages		
Timestamp User 2015-07-09 10:56:00+00 Keith Jolley Corresponding isolate data has a	Message already been submitted (ids 43721-43722).	
	Message: Append	

Your submission will then be listed under 'Pending submissions' on your submission page.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query           Download: Alleles   MLST profiles           Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Logged in: Keith Jolley (keith2). @Log out   Change password	Help 🗹
Manage submissions	
Submit new data	
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.	
Submission type:	
alleles     MLST profiles	
Pending submissions	
You have submitted the following submissions that are pending curation:	
Submission id         Submitted         Updated         Type         Details           BIGSdb_20150709115437_16625_11378         2015-07-09         2015-07-09         profiles         2 MLST profiles	
Return to index page	

# 14.4 Isolate submission

New isolate data can only be submitted from within the appropriate isolate database. You may be required to submit isolate data if you would like to get a new MLST sequence type defined, but this depends on individual database policy.

# 14.4.1 Start

Click the 'isolates' link under submission type on the submission management page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Logged in: Keith Jolley (keith). II+Log out   Change password
Manage submissions
Submit new data
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.
Submission type:
• isolates
Return to index page

Download the Excel submission template.

Query:         Search   Browse   Profile/ST   List           Breakdown:         Isolate fields   Scheme/alleles   Publications           Links:         Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Logged in: Keith Jolley (keith). House out   Change password	
Manage submissions	
Submit new isolates	
Paste in your isolates for addition to the database using the template available below.	
<ul> <li>Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.</li> <li>Enter references for your isolates as a semi-colon (;) separated list of PubMed ids.</li> <li>You can also upload additional allele fields along with the other isolate data - simply create a new column with the loce</li> </ul>	us name.
<ul> <li>Download tab-delimited header for your spreadsheet - use 'Paste Special O Text' to paste the data.</li> <li>Download submission template (xlsx format)</li> </ul>	
Please paste in tab-delimited text (include a field header line)	Action —
	Submit
a.	

# 14.4.2 Paste in isolate data

Fill in the template. Some fields are required and cannot be left blank. Check the 'Description of database fields' link on the database contents page to see a description of the fields and allowed values where these have been defined. Where allowed values have been set, the template will have dropdown boxes (although these require newer versions of Excel to work).

Some databases may have hundreds of loci defined, and most will not have a column in the template. You can add new columns for any loci that have been defined and for which you would like to include allelic information for. These locus names must be the primary locus identifier. A list of loci can be found in the 'allowed\_loci' tab of the Excel submission template.

Copy and paste the entire contents of the submission worksheet. Click submit.

PubMLST Bre	ery: Search   akdown: Isol ks: Contents	ate fields	Scheme/all	eles   Publ		ions   Data	base submi:	ssions		
Logged in: Keith Jolley (keith	h). ເ⇔Log out	Change pass	word							
Manage subm	ission	5								
Submit new isolate	-									
Paste in your isolates for Enter aliases (alt Enter references You can also uplo Download tab-de	ternative nan for your isola oad addition: elimited head	ies) for you ites as a s al allele fie er for your	r isolates a emi-colon (; ids along wi spreadshee	s a semi-c ) separate th the othe	olon (;) sepa d list of PubN r isolate data	irated list. Aed ids. a - simply c			ith the lo	icus name.
Download submi     Please paste in tab-	delimited tex	t (include	a field head							Action
isolate aliases disease source serotype amoxicillin chloramphenicol rifampicin	epidemic sero_sub sulphona	logy type mide henicol	ceftria range	serogr penici xone cefota	oup llin ceftria	MLEE d penici xone ra cefota	age_mtl lesignation llin_ran unge xime_ran loxacin	òn ge ge		Submit
pending_assembl pdhC pgm UK322	y FetA VR	comment	3	abcZ PorA V 2015	adk	aroE	fumC	gdh		
meningitis and B	septicaer	üa	blood		Neisser	ia meni	ngitidis			
2 UK325	3	4 UK	3	8 2015	4	6	F1-5	5	2	
septicaemia B	CSF		Neisser	ia meni	ngitis					
2	3	4	3	8	4	6	F1-5	5-1	2	
									ii.	1

Some basic checks will be performed. These include checking all field values conform to allowed lists or data types. The submission cannot proceed if any checks fail.

ged in: Kei	ith Jolley (keitt	i). 🕩Log out	Change pass	word								
anage	e subm	ission	s									
rror:												
K325 has	problems - s	pecies: 'Nei	isseria mer	ingitis' is no	ot on the list	ofallowed	values for t	his field.				
				-								
ubmit n	ew isolate	5										
aste in you	ur isolates for	addition to	the databas	se using the	template a	vailable bel	ow.					
• Ente	er aliases (alt	ernative nar	nes) for you	r isolates a	s a semi-co	lon (;) sepa	rated list.					
<ul> <li>Ente</li> </ul>	er references	for your isol	ates as a s	emi-colon (;	) separated	list of Publ	led ids.					
• You	can also uplo	ad addition	al allele fiel	ds along wi	th the other	isolate data	a - simply c	reate a new	column wi	th the locu	s name.	
	/nload tab-de				et - use 'Pas	te Special (	Text to pa	iste the data				
	vnload tab-de vnload submi				et - use 'Pas	te Special (	Text to pa	iste the data				
• Dow		ssion temp	late (xlsx for	mat)		te Special (	Text to pa	iste the data				
• Dow Please isolate	vnload submi paste in tab- e aliases	ssion temp delimited te referenc	late (xisx for xt (include a ce <i>s</i>	mat) a field head country	erline) — region	year	age_yr	age_mt)	n sex		Action	
• Dow Please isolate disease	vnload submi paste in tab- e aliases e source	ssion temp delimited te referen epidemic	late (xisx for xt (include a ces plogy	mat) a field head country species	region serogro	year	age_yr MLEE_d	age_mt) esignatio	h sex	_		
Dow     Please     isolate     disease     serotyp	vnload submi paste in tab- e aliases e source pe	ssion temp delimited te referen epidemi sero_sul	late (xlsx for xt (include a ces clogy otype	mat) a field head country species ET_no	region serogro penicil	year oup lin	age_yr MLEE_d penici	age_mth esignatio llin_rang	h sex	_		
Dow     Please     isolate     disease     serotyp     amoxic:	vnload submi paste in tab- e aliases e source	ssion temp delimited te reference epidemic sero_sul sulphon	late (xlsx for xt (include a ces ology btype amide	mat) a field head country species ET_no ceftria	region serogro penicil xone	year pup lin ceftria	age_yr MLEE_d penici axone_ra	age_mth esignatio llin_rang	n sex on ge			
• Dow Please isolate disease serotyp amoxic: chlorar rifamp:	vnload submi paste in tab- e aliases e source pe illin mphenicol icin	ssion temp delimited te referen: epidemic sero_sul sulphon chloramp rifampic	late (xisx for xt (include a ces ology otype amide ohenicol cin_rang	mat) a field heado country species ET_no ceftria _range e	region serogro penicil xone cefotax ciprofl	year Dup lin ceftria time .oxacin	age_yr MLEE_d penici axone_ra cefota ciprof	age_mth esignatio llin_rano nge xime_rano loxacin_r	h sex on ge ge range			
• Dow Please isolate disease serotyp amoxic: chlorar rifamp: pending	nload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembl:	ssion temp delimited te referen epidemid sero_sul sulphon chloram rifampid y	At (include a ces blogy btype amide phenicol cin_rang comment	mat) a field heado country species ET_no ceftria range e s	er line) region serogro penicil xone cefotax ciprofl abcZ	year oup lin ceftria time oxacin adk	age_yr MLEE_d penici xone_ra cefota	age_mtl esignatio llin_rano nge xime_rano	n sex on ge ge			
• Dow Please isolate disease serotyp amoxic: chlorar rifamp:	vnload submi paste in tab- e aliases e source pe illin mphenicol icin	ssion temp delimited te referen epidemid sero_sul sulphon chloram rifampid y	late (xisx for xt (include a ces ology otype amide ohenicol cin_rang	mat) a field heado country species ET_no ceftria range e s	region serogro penicil xone cefotax ciprofl	year oup lin ceftria time oxacin adk	age_yr MLEE_d penici axone_ra cefota ciprof	age_mth esignatio llin_rano nge xime_rano loxacin_r	h sex on ge ge range			
• Dow Please isolate disease serotyp amoxic: chlorar rifamp: pending pdhC UK322 mening:	nload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembl:	delimited te referen epidemid sero_sul sulphon chloramp rifampio Y FetA_VR	At (include a ces ology otype amide phenicol cin_rang comment PorA_VR UK	mat) a field heado country species ET_no ceftria range e s	er line) region serogro penicil xone cefotax ciprofl abcZ PorA_VF	year pup lin ceftria time .oxacin adk 22	age_yr MLEE_d penici ixone_ra cefota ciprof aroE	age_mth esignatio llin_rano nge xime_rano loxacin_r	h sex on ge ge range			
• Dow Please isolate disease serotyp amoxic: chlorar rifamp: pending pdhC UK322	vnload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembl: pgm	delimited te referen epidemid sero_sul sulphon chloramp rifampio Y FetA_VR	At (include a ces ology otype amide phenicol cin_rang comment PorA_VR UK	mat) a field head country species ET_no ceftria range e s 1	er line) region serogro penicil xone cefotax ciprofl abcZ PorA_VF	year pup lin ceftria time .oxacin adk 22	age_yr MLEE_d penici ixone_ra cefota ciprof aroE	age_mtl esignatid llin_rang nge xime_rang loxacin_: fumC	h sex on ge ge range			
• Dow Please isolate disease serotyp amoxic: chlorar rifamp: pending pdhC UK322 mening:	vnload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembl: pgm	delimited te referen epidemid sero_sul sulphon chloramp rifampio Y FetA_VR	At (include a ces ology otype amide phenicol cin_rang comment PorA_VR UK	mat) a field head country species ET_no ceftria range e s 1	er line) region serogro penicil xone cefotax ciprofl abcZ PorA_VF	year pup lin ceftria time .oxacin adk 22	age_yr MLEE_d penici ixone_ra cefota ciprof aroE	age_mtl esignatid llin_rang nge xime_rang loxacin_: fumC	h sex on ge ge range	2		
• Dow Please isolate disease serotyp amoxic: chlorar rifamp: pending pdhC UK322 mening:	vnload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembl pgm itis and i	ssion temp delimited te epidemi, sero_sul sulphon, chloram, rifampi, y FetA_VR septicae: 3	ate (xisx for xt (include a ces ology otype amide phenicol cin_rang comment PorA_VR UK mia	mat) a field head country species ET_no ceftria range e s 1 blood 3	er line) region serogro penicil xone cefotax ciprofl abc2 PorA_VF 2015	year uup lin ceftria time ooxacin adk 12 Neisser 4	age_yr MLEE_d penici ixone_ra cefota ciprof aroE ria meni	age_mtl esignatid llin_ran nge xime_ran loxacin_: fumC ngitidis	h sex on ge ge gah	2		
• Dow Please isolate disease serotyp amoxic: chlorar rifamp; pending pdhC UK322 mening: B	vnload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembl pgm itis and i 2	ssion temp referen epidemid sero_sul sulphon chloram rifampio Y FetA_VR septicae	ate (xisx for xt (include a ces bology obvype amide phenicol pin_rang comment ForA_VR UK mia	mat) a field head country species ET_no ceftria range e s 1 blood 3	er line) region serogro penicil xone cefotax ciprofl abc2 PorA_VF 2015	year uup lin ceftria time ooxacin adk 12 Neisser 4	age_yr MLEE_d penici ixone_ra cefota ciprof aroE ria meni	age_mtl esignatid llin_ran nge xime_ran loxacin_: fumC ngitidis	h sex on ge ge gah	2		
• Dow Please isolate disease serotyn amoxic: chlorar rifamp: pendinG UK322 mening: B UK325 septica	vnload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembl pgm itis and i 2	ssion temp delimited te epidemi, sero_sul sulphon, chloram, rifampi, y FetA_VR septicae: 3	ate (xisx for xt (include a ces bology obvype amide phenicol pin_rang comment ForA_VR UK mia	mat) a field head country species ET_no ceftria range e s 1 blood 3	er line) region serogro penicil xone cefotax ciprofl abc2 PorA_VF 2015	year uup lin ceftria time ooxacin adk 12 Neisser 4	age_yr MLEE_d penici ixone_ra cefota ciprof aroE ria meni	age_mtl esignatid llin_ran nge xime_ran loxacin_: fumC ngitidis	h sex on ge ge gah	2		
• Dow Please isolate disease serotyn amoxic: chlorar rifamp: pendinG UK322 mening: B UK325 septica	vnload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembly pgm itis and : 2 aemia	ssion temp delimited te epidemi, sero_sul sulphon, chloram rifampi, y FetA_VR septicae: 3 CSF	ate (xisx for xt (include a ces blogy btype amide phenicol cin_rang comment PorA_VR mia 4 UK	mat) a field head country species ET_no ceftria range e s 1 blood 3 Neisser	er line) region serogro penicil abc2 PorA_VF 2015 8 2015 ia menir	year uup lin ceftris time ooxacin adk 12 Neisser 4 gitis	age_yr MLEE_d penici xxone_ra cefota ciprof aroE :ia meni	age_mtl esignati llin_ran nge xime_ran fumC fumC ngitidis F1-5	n sex on ge ge gadh	-		
• Dow Please isolate disease serotyn amoxic: chlorar rifamp: pendinG UK322 mening: B UK325 septica	vnload submi paste in tab- e aliases e source pe illin mphenicol icin g_assembly pgm itis and : 2 aemia	ssion temp delimited te epidemi, sero_sul sulphon, chloram rifampi, y FetA_VR septicae: 3 CSF	ate (xisx for xt (include a ces blogy btype amide phenicol cin_rang comment PorA_VR mia 4 UK	mat) a field head country species ET_no ceftria range e s 1 blood 3 Neisser	er line) region serogro penicil abc2 PorA_VF 2015 8 2015 ia menir	year uup lin ceftris time ooxacin adk 12 Neisser 4 gitis	age_yr MLEE_d penici xxone_ra cefota ciprof aroE :ia meni	age_mtl esignati llin_ran nge xime_ran fumC fumC ngitidis F1-5	n sex on ge ge gadh	-		

Provided the checks pass, you will then be able to add additional information to your submission.

# 14.4.3 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Append'.

u are submittin																		
	g the following isolates: Dow	nload <sub>c</sub>	and the second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second se															
olate country		source		serogroup					pdhC									
K322 UK	meningitis and septicaemia septicaemia		Neisseria meningitidis Neisseria meningitidis	B	2	3	4 3		4	6	F1-5 F1-5	5 5-1		2				
E-mail			Action															
	ent to keith.jolley@zoo.ox.ac.u		Finalize submission!															
E-mail submi	ssion undates																	
	solon apaaloo																	
Supporting file																		
Supporting file ease upload ar	s ny supporting files required for	r curatio	n. Ensure that these are	named una	ambigu	iously c	or add a	n expla	inatory	note s	o that the	y can be l	linked to	o the app	ropriate	submiss	ion item.	
ease upload ar		r curatio	n. Ensure that these are	named una	ambigu	iously c	or add a	n expla	inatory	note s	o that the	y can be l	linked to	o the app	ropriate	submiss	ion item.	
ease upload ar dividual filesize	ny supporting files required for is limited to 32 MB.	r curatio ad files	n. Ensure that these are	named una	ambigu	iously c	or add a	n expla	inatory	note s	o that the	y can be l	linked to	o the app	ropriate	submiss	ion item.	
ease upload ar dividual filesize Browse No	ny supporting files required for is limited to 32 MB.		n. Ensure that these are	named una	ambigu	iously c	or add a	n expla	inatory	note s	o that the	y can be l	linked to	o the app	ropriate	submiss	ion item.	
ease upload ar dividual filesize Browse No Messages	ny supporting files required for is limited to 32 MB. files selected. Uploa	ad files	]	named una	ambigu	iously c	or add a	n expla	inatory	note s	o that the	y can be l	linked to	o the app	ropriate	submiss	ion item.	
ease upload ar dividual filesize Browse No Messages	ny supporting files required for is limited to 32 MB.	ad files	]	named una	ambigu	iously c	or add a	n expla	inatory	note s	o that the	y can be l	linked to	o the app	ropriate	submiss	ion item.	

The message will be attached. A curator may respond to the message and attach their own, with the full conversation becoming part of the submission record.

PubML	ST) Bre		fields   Sch	neme/allel	es   Publications es/sequences def	initions   D	)atabase	e subm	nission	IS								
Logged in: Keith	Jolley (keit	h). 🕩 Log out   Cha	nge password	i														Help 🗹
Manage	subm	issions																
Submission	1: BIGSd	lb 201507091	21747 1	342 996	24													
Isolates																		
	mittina the	following isolat	es: Downlo	ad 🕞														
isolate cou UK322 U		disease ingitis and septi		urce ood Neis	species seria meningitidi		up abcž 2	adk 3	aroE 1	fumC 3	jdh pdh 8 4	C pgm 6	FetA_VR F1-5	PorA_VR1 5	PorA_VR2 2			
UK325 U		septicaemia			seria meningitidi		2	3	4	3	18 4	6	F1-5	5-1	2			
— E-mail —				Actio	n													
Updates will	be sent to	o keith.jolley@zo	o.ox.ac.uk.	Final	ize submission!													
🗹 E-mail si	ubmissior	n updates																
	g files —																	
		pporting files rec nited to 32 MB.	uired for cu	iration. En	sure that these ar	e named u	inambig	uously	or add	d an ex	planato	ry note s	so that the	ey can be lin	ked to the ap	opropriate	submissi	on item.
Browse	No files :	selected.	Upload f	iles														
Messages	s											_						
	stamp	User	Lthink LIK?	25 hos ou	new MLST profile	Message	lofinition	of o p	ow CT	numb	ar Thora							
2015-07-09	11.19.12	+00 Keith Jolley	T UNITK OK3	20 nas a	new MLST prome	requiring c	lennuon	oran	ewsi	numb	er. Trian	KS.						
									Mess	age:	Appen	d						

# 14.4.4 Add supporting files

You can add any files required to support the submission. You may, for example, wish to include a genome sequence for an isolate record (contigs in FASTA format). If you are doing this, make sure that the filename can be unambiguously linked to the appropriate isolate record and *add a message*.

Files can be added to the submission by clicking the 'Browse' button in the 'Supporting files' section.

Query: Search   Browse   Profile/ST   List           Breakdown: Isolate fields   Scheme/alleles   Publications           Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions
Logged in: Keith Jolley (keith). 19Log out   Change password
Manage submissions
Submission: BIGSdb_20150709121747_1342_99624
UK322       UK       meningitis and septicaemia       blood       Neisseria meningitidis       B       2       3       4       3       8       4       6       F1-5       5       2         UK325       UK       septicaemia       CSF       Neisseria meningitidis       B       2       3       4       3       18       4       6       F1-5       5-1       2
E-mail     Action       Updates will be sent to keith.jolley@zoo.ox.ac.uk.     Finalize submission!       Image: Second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second second
Supporting files Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB. Browse No files selected. Upload files Messages
Timestamp         User         Message           2015-07-09 11:19:12+00         Keith Jolley   think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.
Message: Append

Select the file in the selection box, then click 'Upload files'.

Query: Search   Browse   Profile/ Breakdown: Isolate fields   Sche Links: Contents   Home   Option	me/alleles   Publications	itions   Databas	e subm	issions						
Logged in: Keith Jolley (keith). Hog out   Change password										Help 🗹
Manage submissions										
Submission: BIGSdb_20150709121747_134	42 99624									
You are submitting the following isolates: Download	d 🕞									
isolate country disease sour UK322 UK meningitis and septicaemia blo		serogroup abc		aroE fumC	gdh pd	hC pgm 4 6	FetA_VR F1-5	PorA_VR1 5	PorA_VR2 2	
	F Neisseria meningitidis	B 2		4 3	18 4	4 6	F1-5	5-1	2	
E-mail	Action									
Updates will be sent to keith.jolley@zoo.ox.ac.uk.	Finalize submission!									
E-mail submission updates										
Supporting files										
Please upload any supporting files required for cura Individual filesize is limited to 32 MB.	ation. Ensure that these are i	named unambig	guously	or add an	explanate	ory note s	o that they	can be link	ked to the ap	propriate submission item.
Browse UK325_genome_contigs.fas Uploa	d files									
Messages										
Timestamp User 2015-07-09 11:19:12+00 Keith Jolley I think UK32		lessage	(	OT	haa Thaa					
2015-07-09 11:19:12+00 Reith Jolley Tthink 0K32	s has a new MLST profile re	quiring delinitio	n or a ne	wsinum	ber. Thar	IKS.				
						.4				
				Message:	Apper	nd				

The file will be uploaded and shown in a table.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Logged in: Keith Jolley (keith). @Log out   Change password	Help 🖉
Manage submissions	
Submission: BIGSdb_20150709121747_1342_99624	
- Isolates	
You are submitting the following isolates: Download cst	
isolate country  disease  source  species  serogroup abcZ adk aroE fumC gdh pdhC pgm FetA_VR PorA_VR1 PorA_VR2	
UK322 UK meningitis and septicaemia blood Neisseria meningitidis B 2 3 4 3 8 4 6 F1-5 5 2	
UK325 UK septicaemia CSF Neisseria meningitidis B 2 3 4 3 18 4 6 F1-5 5-1 2	
E-mail Action	
Updates will be sent to keith.jolley@zoo.ox.ac.uk. Finalize submission!	
C E-mail submission updates	
- Supporting files	
Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.	
Browse_ No files selected. Upload files Filename Size Delete	
UK325_genome_contigs.fas 2 MB	
Delete selected files	
- Messages	
Timestamp User Message	
2015-07-09 11:19:12+00 Keith Jolley I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.	
Message: Append	

Files can be removed from the submission by checking the appropriate 'Delete' box and clicking 'Delete selected files'.

# 14.4.5 Finalize submission

Make sure the 'E-mail submission updates' box is checked if you wish to receive E-mail notification of the result of your submission. This setting is remembered between sessions.

Click 'Finalize submission!'.

Court:         Search   Browse   Profile/ST   List           Breakdown:         Isolate fields   Scheme/alleles   Publications           Links:         Contents   Home   Options   Profiles/sequences definitions   Database submissions															
Logged in: Ke	eith Jolley (keit	th). 🕩 Log out   Change pass	word											He	elp 🕑
Manag	e subm	nissions													
Submiss	ion: BIGSd	lb 20150709121747	1342	99624											
Isolate	es —														
You are s	submitting the	following isolates: Dov	nload	251											
isolato	country	disease	SOURCE		orogroup	abo7	odk	aroE	fum( add	ndb(	Dam	FotA VD	PorA_VR1		
isolate UK322				e species s Neisseria meningitidis	B	2	3	4	3 8	4	6	F1-5	5	2	
UK325	UK	septicaemia	CSF	Neisseria meningitidis	В	2	3	4	3 18	4	6	F1-5	5-1	2	
— E-mail				Action											
Updates	will be sent to	o keith.jolley@zoo.ox.ac	uk.	Finalize submission!											
🗹 E-ma	il submissior	n updates													
	orting files —														
		pporting files required fo on item. Individual filesia		on. Ensure that these are n	amed una	ambigu	uously	y or ad	id an expla	natory	note s	o that the	y can be lini	ced to the	
				10 0 02 MD.											
Browse			ad files												
UK325	Filename genome con	Size Delete													
	selected files														
— Messa															
	imestamp	User		Me	essage										
	2015-07-09 11:19:12+00 Keith Jolley I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.														
								Mes	sage: A	opend					

Your submission will then be listed under 'Pending submissions' on your submission page.

Query: Search   Browse   Profile/ST   List Breakdown: Isolate fields   Scheme/alleles   Publications Links: Contents   Home   Options   Profiles/sequences definitions   Database submissions	
Logged in: Keith Jolley (keith). Hog out   Change password	Help 🖉
Manage submissions	
Submit new data	
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.	
Submission type:	
isolates	
Pending submissions	
You have submitted the following submissions that are pending curation:	
Submission id         Submitted         Updated         Type         Details           BIGSdb_20150709121747_1342_99624         2015-07-09         2015-07-09         isolates         2	
Return to index page	

# 14.5 Removing submissions from your notification list

Once a submission has been closed by a curator, the results will be displayed in your 'Manage submissions' area. You can remove submissions once you have noted the result by clicking the 'Remove' link.

Query: Sequences   Batch sequences   Compare alleles   Profile/ST   Batch profiles   List   Browse   Query Download: Alleles   MLST profiles Links: Contents   Home   PorA   FetA   Options   Isolate Database	
Logged in: Keith Jolley (keith2). @Log out   Change password	Help 🖓
Manage submissions	
Submit new data	
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.	
Submission type:	
alleles     MLST profiles	
Recently closed submissions	
You have submitted the following submissions which are now closed:	
Submission id         Submitted         Updated         Type         Details         Outcome         Remove           BIGSdb_20150714071515_10601_27668         2015-07-14         2015-07-14         alleles         3 NEIS0001 (lpxC) sequences         (b)         (c)         (c)	
Return to index page	

Alternatively, submissions will be removed automatically a specified period of time after closure. By default, this time is 90 days, but this can vary depending on the site configuration.

# **RESTful Application Programming Interface (API)**

The REST API allows third-party applications to retrive data stored within BIGSdb databases or to send new submissions to database curators. To use the REST API, your application will make a HTTP request and parse the response. The response format is JSON (except for routes that request a FASTA or CSV file).

Access to protected resources, i.e. those requiring an account, can be accessed via the API using OAuth authentication.

# 15.1 Passing additional/optional parameters

Optional parameters can be passed as arguments to the query URL by adding a '?' followed by the first argument and its value (separated by a '='). Additional parameters are separated by a '&', e.g.

http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates?page=2&page\_size=100

# **15.2 Resources**

- GET / or /db List site resources
- GET /db/{database} List database resources
- GET /db/{database}/loci List loci
- GET /db/{database}/loci/{locus} Retrieve locus record
- GET /db/{database}/loci/{locus}/alleles Retrieve list of alleles defined for a locus
- GET /db/{database}/loci/{locus}/alleles\_fasta Download alleles in FASTA format
- GET /db/{database}/loci/{locus}/alleles/{allele\_id} Retrieve full allele information
- GET /db/{database}/schemes List schemes
- GET /db/{database}/schemes/{scheme\_id} Retrieve scheme information
- GET /db/{database}/schemes/{scheme\_id}/fields/{field} Retrieve information about scheme field
- GET /db/{database}/schemes/{scheme\_id}/profiles List allelic profiles defined for scheme
- GET /db/{database}/scheme\_id}/profiles\_csv Download allelic profiles in CSV (tab-delimited) format
- GET /db/{database}/schemes/{scheme\_id}/profiles/{profile\_id} Retrieve allelic profile record
- GET /db/{database}/isolates Retrieve list of isolate records

- *GET /db/{database}/isolates/{isolate\_id}* Retrieve isolate record
- GET /db/{database}/isolates/{isolate\_id}/allele\_designations Retrieve list of allele designations
- GET /db/{database}/isolate\_id}/allele\_designations/{locus} Retrieve full allele designation record
- GET /db/{database}/isolates/{isolate\_id}/allele\_ids Retrieve allele identifiers
- *GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele\_designations* Retrieve scheme allele designation records
- *GET /db/{database}/isolate\_id}/schemes/{scheme\_id}/allele\_ids* Retrieve list of scheme allele identifiers
- GET /db/{database}/isolates/{isolate\_id}/contigs Retrieve list of contigs
- GET /db/{database}/isolates/{isolate\_id}/contigs\_fasta Download contigs in FASTA format
- GET /db/{database}/contigs/{contig\_id} Retrieve contig record
- GET /db/{database}/fields Retrieve list of isolate provenance field descriptions
- GET /db/{database}/users/{user\_id} Retrieve user information
- GET /db/{database}/projects Retrieve list of projects
- GET /db/{database}/projects/{project\_id} Retrieve project information
- GET /db/{database}/projects/{project\_id}/isolates Retrieve list of isolates belonging to a project
- GET /db/{database}/submissions Retrieve list of submissions
- POST /db/{database}/submissions Create new submission
- GET /db/{database}/submissions/{submission\_id} Retrieve submission record
- DELETE /db/{database}/submissions/{submission\_id} Delete submission record
- GET /db/{database}/submissions/{submission\_id}/messages Retrieve submission correspondence
- POST /db/{database}/submissions/{submission\_id}/messages Add submission correspondence
- GET /db/{database}/submissions/{submission\_id}/files retrieve list of supporting files uploaded for submission
- POST /db/{database}/submissions/{submission\_id}/files Upload submission supporting file
- GET /db/{database}/submissions/{submission\_id}/files/{filename} Download submission supporting file
- DELETE /db/{database}/submissions/{submission\_id}/files/{filename} Delete submission supporting file

## 15.2.1 GET / or /db - List site resources

## Required route parameters: None

**Optional query parameters:** None

#### Example request URI: http://rest.pubmlst.org/

**Response:** List of resource groupings (ordered by name). Groups may consist of paired databases for sequence definitions and isolate data, or any set of related resources. Each group contains:

- name [string] short name (usually a single word)
- description [string] fuller description
- databases [array] list of database objects, each consists of three key/value pairs:
  - name [string] name of database config

- description [string] short description of resource
- href [string] URI to access resource

# 15.2.2 GET /db/{database} - List database resources

These will vary depending on whether the resource is an isolate or a sequence definition database.

Required route parameter: database [string] - Database configuration name

**Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates

**Response:** Object containing a subset of the following key/value pairs:

- *fields* [string] URI to isolate provenance field information
- · isolates [string] URI to isolate records
- schemes [string] URI to list of schemes
- loci [string] URI to list of loci
- projects [string] URI to list of projects

## 15.2.3 GET /db/{database}/loci - List loci

Required route parameter: database [string] - Database configuration name

#### **Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/loci

Response: Object containing:

- records [int] Number of loci.
- loci [array] List of *URIs to defined locus records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] Some or all of the following:
  - previous URI to previous page of results
  - next URI to next page of results
  - first URI to first page of results
  - last URI to last page of results
  - return\_all URI to page containing all results (paging disabled)

# 15.2.4 GET /db/{database}/loci/{locus} - Retrieve locus record

Provides information about a locus, including links to allele sequences (in seqdef databases).

#### **Required route parameters:**

- database [string] Database configuration name
- locus [string] Locus name

### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/loci/abcZ

Response: Object containing a subset of the following key/value pairs:

- id [string] locus name
- data\_type [string] 'DNA' or 'peptide'
- allele\_id\_format [string] 'integer' or 'text'
- allele\_id\_regex [string] regular expression constraining allele ids
- common\_name [string]
- aliases [array] list of alternative names of the locus
- length\_varies [boolean]
- length [integer] length if alleles are of a fixed length
- coding\_sequence [boolean]
- orf [integer] 1-6
- schemes [array] list of scheme objects, each consisting of:
  - scheme [string] URI to scheme information
  - description [string]
- min\_length [integer] (seqdef databases) minimum length for variable length loci
- max\_length [integer] (seqdef databases) maximum length for variable length loci
- alleles [string] (seqdef databases) URI to list of allele records
- alleles\_fasta [string] (seqdef databases) URI to FASTA file of all alleles of locus
- curators [array] (seqdef databases) list of URIs to user records of curators of the locus
- publications [array] (seqdef databases) list of PubMed id numbers of papers describing the locus
- full\_name [string] (seqdef databases)
- product [string] (seqdef databases)
- description [string] (seqdef databases)
- extended\_attributes [array] (seqdef databases) list of extended attribute objects. Each consists of a subset of the following fields:
  - field [string] field name
  - value\_format [string] 'integer', 'text', or 'boolean'
  - value\_regex [string] regular expression constraining value
  - description [string] description of field

- length [integer] maximum length of field
- required [boolean]
- allowed\_values [array] list of allowed values
- genome\_position [integer] (isolate databases)

# 15.2.5 GET /db/{database}/loci/{locus}/alleles - Retrieve list of alleles defined for a locus

## **Required route parameters:**

- database [string] Database configuration name
- locus [string] Locus name

#### **Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/loci/abcZ/alleles

#### Response: Object containing:

- records [int] Number of alleles
- alleles [array] List of *URIs to defined allele records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] Some or all of the following:
  - previous URI to previous page of results
  - next URI to next page of results
  - first URI to first page of results
  - last URI to last page of results
  - return\_all URI to page containing all results (paging disabled)

# 15.2.6 GET /db/{database}/loci/{locus}/alleles\_fasta - Download alleles in FASTA format

#### **Required route parameters:**

- database [string] Database configuration name
- locus [string] Locus name

#### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/loci/abcZ/alleles\_fasta

Response: FASTA format file of allele sequences

# 15.2.7 GET /db/{database}/loci/{locus}/alleles/{allele\_id} - Retrieve full allele information

## **Required route parameters:**

- database [string] Database configuration name
- locus [string] Locus name
- allele\_id [string] Allele identifier

## **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/loci/abcZ/alleles/5

**Response:** Object containing the following key/value pairs:

- locus [string] URI to locus description
- allele\_id [string] allele identifier
- sequence [string] sequence
- status [string] either 'Sanger trace checked', 'WGS: manual extract', 'WGS: automated extract', or 'unchecked'
- sender [string] URI to user details of sender
- curator [string] URI to user details of curator
- date\_entered [string] record creation date (ISO 8601 format)
- datestamp [string] last updated date (ISO 8601 format)

# 15.2.8 GET /db/{database}/schemes - List schemes

## Required route parameter: database [string] - Database configuration name

## **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/schemes

## **Response:**

- records [integer] Number of schemes
- schemes [array] list of scheme objects, each containing:
  - scheme [string] URI to scheme information
  - description [string]

# 15.2.9 GET /db/{database}/schemes/{scheme\_id} - Retrieve scheme information

Includes links to allelic profiles (in seqdef databases, if appropriate). Required route parameters:

- database [string] Database configuration name
- scheme\_id [integer] Scheme id number

## **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/schemes/1

Response: Object containing a subset of the following key/value pairs:

- id [integer]
- description [string]
- locus\_count [integer] number of loci belonging to scheme
- loci [array] list of URIs to locus descriptions
- has\_primary\_key\_field [boolean]
- fields [array] list of URIs to scheme field descriptions
- primary\_key\_field [string] URI to primary key field description
- profiles [string] URI to list of profile definitions (only seqdef databases)
- profiles\_csv [string] URI to tab-delimited file of all scheme profiles
- curators [array] (seqdef databases) list of URIs to user records of curators of the scheme

# 15.2.10 GET /db/{database}/schemes/{scheme\_id}/fields/{field} - Retrieve information about scheme field

#### **Required route parameters:**

- database [string] Database configuration name
- scheme\_id [integer] Scheme id number
- field [string] Field name

#### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/schemes/1/fields/ST

Response: Object containing the following key/value pairs:

- field [string] field name
- type [string] data type of field (integer or text)
- primary\_key [boolean] true if field is the scheme primary key

# 15.2.11 GET /db/{database}/schemes/{scheme\_id}/profiles - List allelic profiles defined for scheme

#### **Required route parameters:**

- database [string] Database configuration name
- scheme\_id [integer] Scheme id

## **Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

#### Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/schemes/1/profiles

## Response: Object containing:

• records [int] - Number of profiles

- profiles [array] List of URIs to defined profile records. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] Some or all of the following:
  - previous URI to previous page of results
  - next URI to next page of results
  - first URI to first page of results
  - last URI to last page of results
  - return\_all URI to page containing all results (paging disabled)

# 15.2.12 GET /db/{database}/schemes/{scheme\_id}/profiles\_csv - Download allelic profiles in CSV (tab-delimited) format

#### **Required route parameters:**

- database [string] Database configuration name
- scheme\_id [integer] Scheme id

#### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/schemes/1/profiles\_csv

Response: Tab-delimited text file of allelic profiles

# 15.2.13 GET /db/{database}/schemes/{scheme\_id}/profiles/{profile\_id} - Retrieve allelic profile record

#### **Required route parameters:**

- database [string] Database configuration name
- scheme\_id [integer] Scheme id
- profile\_id [string/integer] Profile id

#### **Optional parameters:** None

#### Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/schemes/1/profiles/11

Response: Object containing the following key/value pairs:

- *primary\_key\_term* [string/integer] The field name is the primary key, e.g. ST. The value is the primary key value (primary\_id used as an argument).
- alleles [object] list of URIs to allele descriptions
- *other\_scheme\_fields* [string/integer] Each scheme field will have its own value if defined. The field name is the name of the field.
- sender [string] URI to user details of sender
- curator [string] URI to user details of curator
- date\_entered [string] record creation date (ISO 8601 format)
- datestamp [string] last updated date (ISO 8601 format)

# 15.2.14 GET /db/{database}/isolates - Retrieve list of isolate records

Required route parameter: database [string] - Database configuration name

### **Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates

#### Response: Object containing:

- records [int] Number of isolates
- isolates [array] List of URIs to isolate records. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] Some or all of the following:
  - previous URI to previous page of results
  - next URI to next page of results
  - first URI to first page of results
  - last URI to last page of results
  - return\_all URI to page containing all results (paging disabled)

# 15.2.15 GET /db/{database}/isolates/{isolate\_id} - Retrieve isolate record

#### **Required route parameters:**

- database [string] Database configuration name
- isolate\_id [integer] Isolate identifier

#### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates/1

Response: Object containing some or all of the following key/value pairs:

- provenance [object] set of key/value pairs. Keys are defined by calling the */fields route*. The fields will vary by database but will always contain the following:
  - id [integer]
  - sender [string] URI to user details of sender
  - curator [string] URI to user details of curator
  - date\_entered [string] record creation date (ISO 8601 format)
  - datestamp [string] last updated date (ISO 8601 format)
- publications [array] (seqdef databases) list of PubMed id numbers of papers that refer to the isolate
- sequence\_bin [object] consists of the following key/value pairs:
  - contigs\_fasta [string] URI to FASTA file containing all the contigs belonging to this isolate
  - contigs [string] URI to list of contig records

- contig\_count [integer] number of contigs
- total\_length [integer] total length of contigs
- allele\_designations [object] consists of the following key/value pairs:
  - allele\_ids URI to list of all allele\_id values defined for the isolate
  - designation\_count number of allele designations defined for the isolate
  - full\_designations URI to list of full allele designation records
- schemes [array] list of scheme objects, each containing the following:
  - description [string] description of scheme
  - loci\_designated\_count [integer] number of loci within scheme that have an allele designated for this isolate.
  - allele\_ids [string] URI to list of all allele\_id values defined for this scheme for this isolate
  - full\_designations [string] URI to list of full allele designation records for this isolate
  - fields [object] consisting of key/value pairs where the key is the name of each scheme field
- projects [array] list of project objects, each containing the following:
  - id [string] URI to project information
  - description [string] description of project
- new\_version [string] URI to newer version of record
- · old\_version [string] URI to older version of record

# 15.2.16 GET /db/{database}/isolates/{isolate\_id}/allele\_designations - Retrieve list of allele designation records

#### **Required route parameters:**

- database [string] Database configuration name
- isolate\_id [integer] Isolate identifier

#### **Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates/1/allele\_designations

#### Response: Object containing:

- records [int] Number of allele designations
- allele\_designations [array] List of *URIs to allele designation records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] Some or all of the following:
  - previous URI to previous page of results
  - next URI to next page of results
  - first URI to first page of results

- last URI to last page of results
- return\_all URI to page containing all results (paging disabled)

# 15.2.17 GET /db/{database}/isolates/{isolate\_id}/allele\_designations/{locus} - Retrieve full allele designation record

#### **Required route parameters:**

- database [string] Database configuration name
- isolate\_id [integer] Isolate identifier
- locus [string] Locus name

## **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates/1/allele\_designations/BACT000065

Response: List of allele\_designation objects (there may be multiple designations for the same locus), each containing:

- locus [string] URI to locus description
- allele\_id [string]
- method [string] either 'manual' or 'automatic'
- status [string] either 'confirmed' or 'provisional'
- comments [string]
- sender [string] URI to user details of sender
- curator [string] URI to user details of curator
- datestamp [string] last updated date (ISO 8601 format)

# 15.2.18 GET /db/{database}/isolates/{isolate\_id}/allele\_ids - Retrieve allele identifiers

#### **Required route parameters:**

- database [string] Database configuration name
- isolate\_id [integer] Isolate identifier

#### **Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates/1/allele\_ids

#### Response: Object containing:

- records [int] Number of allele id objects
- allele\_ids [array] List of allele id objects, each consisting of a key/value pair where the key is the locus name. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] Some or all of the following:

- previous URI to previous page of results
- next URI to next page of results
- first URI to first page of results
- last URI to last page of results
- return\_all URI to page containing all results (paging disabled)

# 15.2.19 GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele\_designations - Retrieve scheme allele designation records

#### **Required route parameters:**

- database [string] Database configuration name
- isolate\_id [integer] Isolate identifier
- · scheme\_id [integer] Scheme identifier

#### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates/1/schemes/1/allele\_designations

## **Response:**

- records [int] Number of allele designation objects
- allele\_designations [array] List of *allele designation objects* for each locus in the specified scheme that has been designated.

# 15.2.20 GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele\_ids -Retrieve list of scheme allele identifiers

#### **Required route parameters:**

- database [string] Database configuration name
- isolate\_id [integer] Isolate identifier
- scheme\_id [integer] Scheme identifier

#### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates/1/schemes/1/allele\_ids

#### **Response:**

- records [int] Number of allele id objects
- allele\_ids [array] List containing allele id objects for each locus in the specified scheme that has been designated. Each allele\_id object contains a key which is the name of the locus with a value that may be either a string, integer or array of strings or integers (required where there are multiple designations for a locus). The data type depends on the allele\_id\_format set for the specific locus.

## 15.2.21 GET /db/{database}/isolates/{isolate\_id}/contigs - Retrieve list of contigs

#### **Required route parameters:**

• database [string] - Database configuration name

• isolate\_id [integer] - Isolate identifier

## **Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates/1/contigs

## Response: Object containing:

- records [int] Number of contigs
- contigs [array] List of *URIs to contig records* Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] Some or all of the following:
  - previous URI to previous page of results
  - next URI to next page of results
  - first URI to first page of results
  - last URI to last page of results
  - return\_all URI to page containing all results (paging disabled)

# 15.2.22 GET /db/{database}/isolates/{isolate\_id}/contigs\_fasta - Download contigs in FASTA format

## **Required route parameters:**

- database [string] Database configuration name
- isolate\_id [integer] Isolate identifier

## **Optional parameter:**

• header [string] - either 'original\_designation' or 'id' (default is 'id'). This selects whether the FASTA header lines contain the originally uploaded FASTA headers or the sequence bin id numbers.

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/isolates/1/contigs\_fasta?header=original\_designation

Response: FASTA format file of isolate contig sequences

# 15.2.23 GET /db/{database}/contigs/{contig\_id} - Retrieve contig record

## **Required route parameters:**

- database [string] Database configuration name
- contig\_id [integer] Contig identifier

## **Optional parameters:** None

## Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/contigs/180062

Response: Contig object consisting of the following key/value pairs:

- id [integer] contig identifier
- isolate\_id [integer] isolate identifier

- sequence [string] contig sequence
- length [integer] length of contig sequence
- method [string] sequencing method
- sender [string] URI to user details of sender
- curator [string] URI to user details of curator
- date\_entered [string] record creation date (ISO 8601 format)
- datestamp [string] last updated date (ISO 8601 format)

# 15.2.24 GET /db/{database}/fields - Retrieve list of isolate provenance field descriptions

## **Required route parameters:**

• database [string] - Database configuration name

## **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/fields

Response: Array of field objects, each consisting of some or all of the following key/value pairs:

- name [string] name of field
- type [string] data type (int, text, date, float)
- length [integer] maximum length of field
- required [boolean] true if field value is required
- min [integer] minimum value for integer values
- max [integer] maximum value for integer values
- regex [string] regular expression that constrains the allowed value of the field
- comments [string]
- allowed values [array] list of allowed values for the field [string]

# 15.2.25 GET /db/{database}/users/{user\_id} - Retrieve user information

Users may be data submitters or curators.

## **Required route parameters:**

- database [string] Database configuration name
- user\_id [integer] User id number

## **Optional parameters:** None

## Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/users/2

Response: Object containing the following key/value pairs:

- id [integer] user id number
- first\_name [string]
- surname [string]

- affiliation [string] institutional affiliation
- email [string] E-mail address

# 15.2.26 GET /db/{database}/projects - Retrieve list of projects

Required route parameter: database [string] - Database configuration name

**Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/projects

## **Response:**

- projects [array] List of project objects, each containing:
  - project [string] URI to project information
  - description [string]
  - isolate\_count [integer] number of isolates in project

# 15.2.27 GET /db/{database}/projects/{project\_id} - Retrieve project information

#### **Required route parameters:**

- database [string] Database configuration name
- project\_id [integer] Project id number

### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/projects/3

Response: Object containing a subset of the following key/value pairs:

- id [integer]
- description [string]
- isolates [string] URI to list of URIs of member isolate records.

# 15.2.28 GET /db/{database}/projects/{project\_id}/isolates - Retrieve list of isolates belonging to a project

### **Required route parameter:**

- database [string] Database configuration name
- project\_id [integer] Project id number

## **Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

#### Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/projects/3/isolates

## Response: Object containing:

· records [int] - Number of isolates in the project

- isolates [array] List of URIs to isolate records. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] Some or all of the following:
  - previous URI to previous page of results
  - next URI to next page of results
  - first URI to first page of results
  - last URI to last page of results
  - return\_all URI to page containing all results (paging disabled)

# 15.2.29 GET /db/{database}/submissions - retrieve list of submissions

Required route parameter: database [string] - Database configuration name

#### **Optional parameters:**

- type [string] either 'alleles', 'profiles' or 'isolates'
- status [string] either 'closed' or 'pending'
- page [integer]
- page\_size [integer]
- return\_all [integer] Set to non-zero value to disable paging.

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_isolates/submissions

#### Response: Object containing:

- records [int] Number of submissions
- submissions [array] List of URIs to submission records
- paging [object] Some or all of the following:
  - previous URI to previous page of results
  - next URI to next page of results
  - first URI to first page of results
  - last URI to last page of results
  - return\_all URI to page containing all results (paging disabled)

# 15.2.30 POST /db/{database}/submissions - create new submission

#### Required route parameter: database [string] - Database configuration name

#### **Required additional parameters:**

- type [string] either:
  - alleles (sequence definition databases only)
  - profiles (sequence definition databases only)
  - isolates (isolate databases only)

The following are required with the specified database type:

#### Allele submissions

- locus [string] name of locus
- technology [string] name of sequencing technology: either '454', 'Illumina', 'Ion Torrent', 'PacBio', 'Oxford Nanopore', 'Sanger', 'Solexa', 'SOLiD', or 'other'
- read\_length [string] read length of sequencing: either '<100', '100-199', '200-299', '300-499', '>500', or any positive integer (only required for Illumina)
- coverage [string] mean coverage of sequencing: either '<20x', '20-49x', '50-99x', '>100x', or any positive integer (only required for Illumina)
- assembly [string] assembly method: either 'de novo' or 'mapped'
- software [string] name of assembly software
- sequences [string] either single raw sequence or multiple sequences in FASTA format

#### **Profile submissions**

- scheme\_id [int] scheme id number
- profiles [string] tab-delimited profile data this should include a header line containing the name of each locus

#### **Isolate submissions**

• isolates [string] - tab-delimited isolate data - this should include a header line containing each field or locus included

#### **Optional parameters:**

- message [string] correspondence to the curator
- email [int] set to 1 to enable E-mail updates (E-mails will be sent to the registered user account address).

#### Response: Object containing:

• submission - URI to submission record

# 15.2.31 GET /db/{database}/submissions/{submission\_id} - Retrieve submission record

#### **Required route parameters:**

- database [string] Database configuration name
- submission\_id [string] Submission id

#### **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/submissions/BIGSdb\_20151013081836\_14559\_14740

Response: Object containing some of the following:

- id [string] Submission id
- type [string] Either 'alleles', 'profiles', 'isolates'
- date\_submitted [string] Submission date (ISO 8601 format)
- datestamp [string] Last updated date (ISO 8601 format)
- submitter [string] URI to user details of submitter

- curator [string] URI to user details of curator
- status [string] either 'started', 'pending', or 'closed'
- outcome [string] either 'good' (data uploaded), 'bad' (data rejected), or 'mixed' (parts of submission accepted)
- correspondence [array] List of correspondence objects in time order. Each contains:
  - user [string] URI to user details of user
  - timestamp [string]
  - message [string]

#### Allele submissions

- · locus [string] name of locus
- technology [string] name of sequencing technology: either '454', 'Illumina', 'Ion Torrent', 'PacBio', 'Oxford Nanopore', 'Sanger', 'Solexa', 'SOLiD', or 'other'
- read\_length [string] read length of sequencing: either '<100', '100-199', '200-299', '300-499', '>500', or any positive integer (only required for Illumina)
- coverage [string] mean coverage of sequencing: either '<20x', '20-49x', '50-99x', '>100x', or any positive integer (only required for Illumina)
- assembly [string] assembly method: either 'de novo' or 'mapped'
- software [string] name of assembly software
- seqs [array] List of sequence objects each containing:
  - seq\_id [string] Sequence identifier
  - assigned\_id [string] Allele identifier if uploaded to the database (otherwise undefined)
  - status [string] Either 'pending', 'assigned', or 'rejected'
  - sequence [string]

#### **Profile submissions**

- scheme [string] URI to scheme information
- profiles [array] List of profile record objects. Each contains:
  - profile\_id [string] Record identifier
  - assigned\_id [string] Profile identifier if uploaded to the database (otherwise undefined)
  - status [string] Either 'pending', 'assigned', or 'rejected'
  - designations [object] containing key/value pairs for each locus containing the allele identifier

#### **Isolate submissions**

isolates [array] - List of isolate record objects. Each contains key/value pairs for included fields.

# 15.2.32 DELETE /db/{database}/submissions/{submission\_id} - Delete submission record

You must be the owner and the record must be closed.

#### **Required route parameters:**

database [string] - Database configuration name

• submission\_id [string] - Submission id

## **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/submissions/BIGSdb\_20151013081836\_14559\_14740 Response: message [string] - 'Submission deleted.'

# 15.2.33 GET /db/{database}/submissions/{submission\_id}/messages - Retrieve submission correspondence

## **Required route parameters:**

- database [string] Database configuration name
- submission\_id [string] Submission id

## **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/submissions/BIGSdb\_20151013081836\_14559\_14740/me.

Response: Array of correspondence objects in time order. Each contains:

- user [string] URI to user details of user
- timestamp [string]
- message [string]

# 15.2.34 POST /db/{database}/submissions/{submission\_id}/messages - Add submission correspondence

## **Required route parameters:**

- database [string] Database configuration name
- submission\_id [string] Submission id

## **Required additional parameter:**

• message [string] - Message text

## **Optional parameters:** None

Response: message [string] - 'Message added.'

# 15.2.35 GET /db/{database}/submissions/{submission\_id}/files - Retrieve list of supporting files uploaded for submission

## **Required route parameters:**

- database [string] Database configuration name
- submission\_id [string] Submission id

## **Optional parameters:** None

Example request URI: http://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/submissions/BIGSdb\_20151013081836\_14559\_14740/file

**Response:** Array of URIs to files

# 15.2.36 POST /db/{database}/submissions/{submission\_id}/files - Upload submission supporting file

## **Required route parameters:**

- database [string] Database configuration name
- submission\_id [string] Submission id

### **Required additional parameters:**

- filename [string] Name of file to store within submission
- upload [base64 encoded data] Raw file data

#### **Optional parameters:** None

Response: message [string] - 'File uploaded.'

# 15.2.37 GET /db/{database}/submissions/{submission\_id}/files/{filename} - Download submission supporting file

#### **Required route parameters:**

- database [string] Database configuration name
- submission\_id [string] Submission id
- filename [string] Name of file

#### **Optional parameters:** None

Response: File download

# 15.2.38 DELETE /db/{database}/submissions/{submission\_id}/files/{filename} Delete submission supporting file

## **Required route parameters:**

- database [string] Database configuration name
- submission\_id [string] Submission id
- filename [string] Name of file

## **Optional parameters:** None

Response: message [string] - 'File deleted.'

# **15.3 Authentication**

Protected resources, i.e. those requiring a user to log in, can be accessed via the API using OAuth (1.0A) authentication (see IETF RFC5849 for details). Third-party client software has to be registered with the BIGSdb site before they can access authenticated resources. The overall three-legged flow works as follows:

- 1. Developer signs up and gets a consumer key and consumer secret specific to their application.
- 2. Application gets a request token and directs user to authorization page on BIGSdb.

- 3. BIGSdb *asks user for authorization* for application to access specific resource using their credentials. A verifier code is provided.
- 4. The application exchanges the request token and OAuth verifier code for an *access token and secret* (these do not expire but may be revoked by the user or site admin).
- 5. Application uses access token/secret to request session token (this is valid for 12 hours).
- 6. All calls to access protected resources are signed using the session token/secret and consumer key/secret.

It is recommended that application developers use an OAuth library to generate and sign requests.

## 15.3.1 Developer sign up to get a consumer key

Application developers should apply to the site administrator of the site running BIGSdb. The administrator can *generate a key and secret* using a script - both of these will need to be used by the application to sign requests.

The client id is usually a 24 character alphanumeric string. The secret is usually a 42 character alphanumeric (including punctuation) string, e.g.

- client\_id: efKXmqp2D0EBlMBkZaGC2lPf
- client\_secret: F\$M)\_+fQ2AFFB2YBDfF9fpHF^qSWJdmmN%L4Fxf5Gur3

## 15.3.2 Getting a request token

- Relative URL: /db/{database}/oauth/get\_request\_token
- Supported method: GET

The application uses the consumer key to obtain a request token. The request token is a temporary token used to initiate user authorization for the application and will expire in 60 minutes. The request needs to contain the following parameters and to be signed using the consumer secret:

- oauth\_consumer\_key
- oauth\_request\_method ('GET')
- oauth\_request\_url (request URL)
- oauth\_signature\_method ('HMAC-SHA1')
- oauth\_signature
- oauth\_timestamp (UNIX timestamp seconds since Jan 1 1970) this must be within 600 seconds of the current time.
- oauth\_callback ('oob' for desktop applications)
- oauth\_nonce (random string)
- oauth\_version ('1.0')

If the application has been registered and has been granted permission to access the specific resource, a JSON response will be returned containing the following parameters:

- oauth\_token
  - This is the request token. It is usually a 32 character alphanumeric string.

- e.g. fKFm0WNhCfbEX8zQm6qhDA8K23FOWDGE

oauth\_token\_secret

- This is the secret associated with the request token. It is usually a 32 character alphanumeric string.
- e.g. aZ0fncP7i5w5jlebdK5zyQ4vrRRVcdnv
- oauth\_callback\_confirmed
  - This parameter is always set to true.

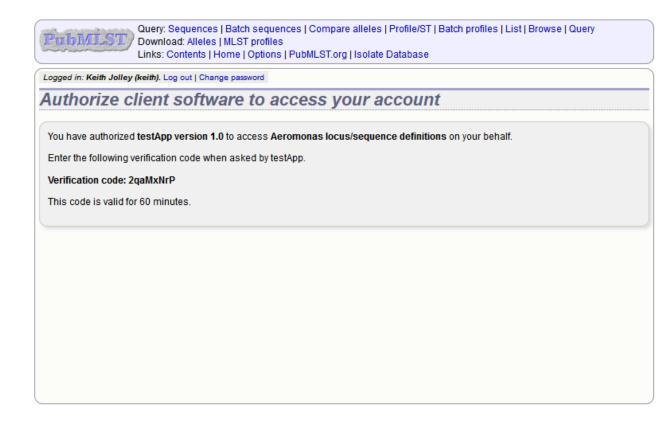
# 15.3.3 Getting user authorization

Once a request token has been obtained, this can be used by the end user to grant permission to access a specific resource to the application. The application should direct the user to the client authorization page (authorizeClient) specific to a database within BIGSdb, e.g. http://pubmlst.org/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst\_neisseria\_seqdef&page=authorizeClient&oauth\_token=fKFm0WNhCfbEX8zQm6qhDA8K23FOWI

The user will be asked if they wish to grant access to the application on their behalf:

PTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT						
Logged in: Keith Jolley (keith). Log out   Change password	Logged in: Keith Jolley (keith). Log out   Change password					
Authorize client software to access your account						
Do you wish for the following application to access data on your beha	alf?					
Application Resource						
testApp version 1.0 Aeromonas locus/sequence definitions	Cancel Authorize					
You will be able to revoke access for this application at any time.						
	)					

If they authorize the access, they will be presented with a verifier code. This should be entered in to the client application which will use this together with the request token to request an access token.



The verifier code is valid for 60 minutes.

#### 15.3.4 Getting an access token

- Relative URL: /db/{database}/oauth/get\_access\_token
- Supported method: GET

The application uses the request token, verifier code and its consumer key to obtain an access token. The access token does not expire but can be revoked by both the end user or the site administrator. The request needs to contain the following parameters and to be signed using the consumer secret and request token secret:

- oauth\_consumer\_key
- oauth\_request\_method ('GET')
- oauth\_request\_url (request URL)
- oauth\_signature\_method ('HMAC-SHA1')
- oauth\_signature
- oauth\_token (request token)
- oauth\_timestamp (UNIX timestamp seconds since Jan 1 1970) this must be within 600 seconds of the current time.
- oauth\_nonce (random string)
- oauth\_version ('1.0')

If the application has been registered and has been granted permission to access the specific resource, a JSON response will be returned containing the following parameters:

- oauth\_token
  - This is the access token. It is usually a 32 character alphanumeric string.
  - e.g. SDrC74ZV15SYSqY8IWZqrRxnyDnNGVFO
- oauth\_token\_secret
  - This is the secret associated with the access token. It is usually a 32 character alphanumeric string.
  - e.g. tYI2SPzgiO02IRVzW4JR1ez6Vvm4gVyv

#### 15.3.5 Getting a session token

- Relative URL: /db/{database}/oauth/get\_session\_token
- Supported method: GET

The application uses the access token and its consumer key to obtain a session token. The session token is valid for 12 hours before it expires. The request needs to contain the following parameters and to be signed using the consumer secret and access token secret:

- oauth\_consumer\_key
- oauth\_request\_method ('GET')
- oauth\_request\_url (request URL)
- oauth\_signature\_method ('HMAC-SHA1')
- oauth\_signature
- oauth\_token (access token)
- oauth\_timestamp (UNIX timestamp seconds since Jan 1 1970) this must be within 600 seconds of the current time.
- oauth\_nonce (random string)
- oauth\_version ('1.0')

If the application has been registered and has been granted permission to access the specific resource, a JSON response will be returned containing the following parameters:

- oauth\_token
  - This is the session token. It is usually a 32 character alphanumeric string.
  - e.g. H8CjIS8Ikq6hwCUqUfF1l4pTaCYl8Ljw
- oauth\_token\_secret
  - This is the secret associated with the session token. It is usually a 32 character alphanumeric string.
  - e.g. RfponbaNPO7tkZ2miHFISk0pMndePNfJ

#### 15.3.6 Accessing protected resources

• Supported method: GET

The application uses the session token and its consumer key to access a protected resource. The request needs to contain the following parameters and to be signed using the consumer secret and session token secret:

• oauth\_consumer\_key

- oauth\_request\_method ('GET')
- oauth\_request\_url (request URL)
- oauth\_signature\_method ('HMAC-SHA1')
- oauth\_signature
- oauth\_token (session token)
- oauth\_timestamp (UNIX timestamp seconds since Jan 1 1970) this must be within 600 seconds of the current time.
- oauth\_nonce (random string)
- oauth\_version ('1.0')

## Frequently asked questions (FAQs)

#### 16.1 General

#### 1. What is the minimum specification of hardware required to run BIGSdb?

The software will run on very modest hardware - a number of PubMLST mirrors have been set up on virtual machines with 1 processor core and 4 GB RAM. This should be considered an absolute minimum specification though. For an installation with only local users, the following minimum is recommended:

- 4 processor cores
- 16 GB RAM
- 50 GB partition for temporary files
- 100 GB partition for databases

As usual, the more RAM that is available the better. Ideally you would want enough RAM that the whole database(s) can reside in memory (an approximation is roughly twice the total size of your contigs), although this is not absolutely required.

Offline jobs, such as *Genome Comparator* will use a processor core each, so if you want to run multiple jobs in parallel then you may want more cores (and memory). Tagging of new genomes using the offline *autotagger* can be run in multi-threaded mode so the more cores available the faster this will be.

As a comparison, the PubMLST site is run on two machines - separate web and database servers. All offline jobs and tagging of genomes is performed on the database server. These have the following specification:

- web server: 16 cores, 64GB RAM
- database server: 64 cores, 1TB RAM, 3TB RAID 10 local storage

## 16.2 Installation

1. BIGSdb is accumulating files in various temp directories - is this normal and how do I clean them out?

See: Periodically delete temporary files.

#### 2. BIGSdb is complaining of an invalid script path - what does this mean?

In your database config.xml file system tag are two attributes - script\_path\_includes and curate\_path\_includes. These contain regexes that the web url to your script (bigsdb.pl and bigscurate.pl respectively) must match. This prevents somebody from accessing a private database using an instance of bigsdb.pl that is not in a protected directory if you're using apache authentication.

So, if you access the script from http://localhost/cgi-bin/bigsdb/bigsdb.pl then you can set script\_path\_includes to something like "/bigsdb/" (which is the default), or "/cgi-bin/" or just "/" if you don't care about this check.

### 16.3 Administration

#### 1. How can I make some isolates public but not others?

The easiest way to do this is to set up two or more separate configuration directories that refer to the database. The URLs to access these will differ by the value of the 'db' attribute, which refers to the name of the configuration directory (in /etc/bigsdb/dbases/). The database view accessed by each of these configurations can be different as can the access restrictions.

#### Example:

We have a database 'bigsdb\_test' that contains data, only some of which we wish to make publicly available. The isolates to make public are all members of a project. First we can make a view of the isolates table that contains only isolates within this project.

For isolates in project id 3, create a database view by logging in to psql as the postgresql user. We will name this view 'public'.:

```
sudo su postgres
psql bigsdb_test
CREATE VIEW public AS SELECT * FROM isolates WHERE id IN (SELECT isolate_id
FROM project_members WHERE project_id=3);
GRANT SELECT ON public TO apache;
```

Create a private configuration that can access everything in the database in /etc/bigsdb/dbases/test\_private. This will be accessible from http://IP\_ADDRESS/cgi-bin/bigsdb/bigsdb.pl?db=test\_private.

The important attributes to set in the system tag of the config.xml file in this directory are::

```
view="isolates"
read_access="authenticated_users"
```

This means that anyone with an account can log in and view all the isolates (because the view is set to the isolates table).

Now create a public configuration in /etc/bigsdb/dbases/test\_public. This will be accessible from http://IP\_ADDRESS/cgi-bin/bigsdb/bigsdb.pl?db=test\_public. It is better to create a symlink to the private config.xml and then override the attributes that are different. So create a symlink to the private config file:

```
cd /etc/bigsdb/dbases/test_public
sudo ln -s ../test_private/config.xml .
```

You can now override the view and access settings. Within /etc/bigsdb/dbases/test\_public, create a file called system.overrides and add the following:

```
view="public"
read_access="public"
```

See also Restricting particular configurations to specific user accounts.

### Appendix

### 17.1 Query operators

Various query forms have operators for use with field values. Available operators are:

• =

- Exact match (case-insensitive).
- contains
  - Match to a partial string (case-insensitive), e.g. searching for clonal complex 'contains' st-11 would return all STs belonging to the ST-11 complex.
- · starts with
  - Match to values that start with the search term (case-insensitive).
- ends with
  - Match to values that end with the search term (case-sensitive).

• >

- Greater than the search term.

• <

- Less than the search term.
- NOT
  - Match to values that do not equal the search term (case-insensitive).
- NOT contain
  - Match to values that do not contain the search term (case-insensitive).

### 17.2 Sequence tag flags

Sequences tagged in the sequence bin can have features indicated by specific flags. The presence of these flags can be queried. These are a superset of *flags available for allele sequences*. Available flags are:

- ambiguous read
  - Genome sequence contains ambiguous nucleotides in coding sequence.

- apparent misassembly
  - Sequence has a region of very high identity to existing allele in one region but looks completely different in another.
- atypical
  - Catch-all term for a sequence that is unusual compared to other alleles of locus.
- · contains IS element
  - Coding sequence is interrupted by insertion sequence.
- · downstream fusion
  - No stop codon present resulting in translation continuing.
- frameshift
  - Frameshift in sequence relative to other alleles, not resulting in internal stop codon.
- internal stop codon
  - Frameshift in sequence relative to other alleles, resulting in internal stop codon.
- no start codon
  - No apparent start codon in immediate vicinity of usual start.
- phase variable: off
  - Coding sequence has a homopolymeric run with a frameshift resulting in a stop codon preventing complete translation.
- truncated
  - Coding sequence is unusually short resulting in a truncated protein (not the same as running off the end of a contig).
- upstream fusion
  - No apparent start codon in immediate vicinity of usual start, likely due to a gene fusion (sequence is transcribed together with upstream coding sequence).

## 17.3 Allele sequence flags

Sequences can be flagged with specific attributes - these are searchable when doing a sequence attribute query. These are a subset of *flags available for tagged sequences*. These are mainly for use with whole genome MLST type data. Multiple flags can be selected by Ctrl-clicking the list. Available flags are:

- atypical
  - Catch-all term for a sequence that is unusual compared to other alleles of locus.
- · contains IS element
  - Coding sequence is interrupted by insertion sequence.
- downstream fusion
  - No stop codon present resulting in translation continuing.
- frameshift
  - Frameshift in sequence relative to other alleles, not resulting in internal stop codon.

- internal stop codon
  - Frameshift in sequence relative to other alleles, resulting in internal stop codon.
- · no start codon
  - No apparent start codon in immediate vicinity of usual start.
- phase variable: off
  - Coding sequence has a homopolymeric run with a frameshift resulting in a stop codon preventing complete translation.
- truncated
  - Coding sequence is unusually short resulting in a truncated protein (not the same as running off the end of a contig).
- upstream fusion
  - No apparent start codon in immediate vicinity of usual start, likely due to a gene fusion (sequence is transcribed together with upstream coding sequence).

**Database schema** 

- Sequence definition database
- Isolate database

#### Index

### A

access control lists, 36 restricting, 36 adding locus, 45, 48, 49, 55, 79 MLST scheme, 78 schemes, 60 allele breakdown, 258 allele definition records, 191 allele designations count, 220 query, 219 status, 222 allele sequence identify, 195 allele sequences adding, 104 alleles, 3 list query, 203 API authentication access token, 355 accessing protected resources, 356 consumer key, 353 request token, 353 session token, 356 user authorization, 354 API resources add submission correspondence, 351 create new submission, 348 DELETE /db/{database}/submissions/{submission id}, 350 DELETE /db/{database}/submissions/{submission\_id}/files/{filename} 352 delete submission record, 350 delete submission supporting file, 352 download alleles in FASTA format, 337

download allelic profiles in CSV (tab-delimited) format. 340 download contigs in FASTA format, 345 download submission supporting file, 352 GET /. 334 GET /db, 334 GET /db/{database}, 335 GET /db/{database}/contigs/{contig\_id}, 345 GET /db/{database}/fields, 346 GET /db/{database}/isolates, 340 GET /db/{database}/isolates/{isolate\_id}, 341 GET/db/{database}/isolates/{isolate\_id}/allele\_designations, 342 GET/db/{database}/isolates/{isolate\_id}/allele\_designations/{locus}, 343 GET /db/{database}/isolates/{isolate\_id}/allele\_ids, 343 GET /db/{database}/isolates/{isolate id}/contigs, 344 GET/db/{database}/isolates/{isolate id}/contigs fasta, 345 GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele 344 GET/db/{database}/isolates/{isolate id}/schemes/{scheme id}/allele 344 GET /db/{database}/loci, 335 GET /db/{database}/loci/{locus}, 335 GET /db/{database}/loci/{locus}/alleles, 337 GET/db/{database}/loci/{locus}/alleles/{allele\_id}, 337 GET /db/{database}/loci/{locus}/alleles\_fasta, 337 GET /db/{database}/projects, 347 GET /db/{database}/projects/{project\_id}, 347 GET /db/{database}/projects/{project\_id}/isolates, GET /db/{database}/schemes, 338 GET /db/{database}/schemes/{scheme\_id}, 338 GET/db/{database}/schemes/{scheme id}/fields/{field}, 339 GET /db/{database}/schemes/{scheme\_id}/profiles, 339

GET/db/{database}/schemes/{scheme\_id}/profiles/{profiles/db, 179 340 GET/db/{database}/schemes/{scheme id}/profiles cs 340 **BLAST**, 275 GET /db/{database}/submissions, 348 breakdown GET /db/{database}/submissions/{submission id}, allele. 258 349 provenance field, 251 GET/db/{database}/submissions/{submission id}/files, scheme, 258 351 sequence bin, 264 GET/db/{database}/submissions/{submission\_id}/files/{filenamedid, 255 352 browse GET/db/{database}/submissions/{submission\_id}/messagescheme profiles, 205 351 **BURST, 279** GET /db/{database}/users/{user\_id}, 346 С list allelic profiles defined for scheme, 339 list database resources, 335 caching list loci, 335 schemes, 39 list schemes, 338 client authorization list site resources, 334 RESTful interface, 101 POST /db/{database}/submissions, 348 client databases, 70 POST/db/{database}/submissions/{submission\_id}/files odon usage, 283 351 Locus Explorer, 249 POST/db/{database}/submissions/{submission\_id}/messages composite fields, 90 351 configuration retrieve allele identifiers, 343 export, 99 retrieve contig record, 345 configuration settings retrieve full allele designation record, 343 validation, 98 retrieve full allele information, 337 count retrieve information about scheme field, 339 allele designations, 220 retrieve isolate record, 341 sequence tags, 222 retrieve list of allele designations, 342 retrieve list of alleles defined for a locus, 337 Ε retrieve list of contigs, 344 export retrieve list of isolate provenance field descriptions, configuration, 99 346 extended attributes retrieve list of isolate records, 340 locus, 57 retrieve list of isolates belonging to a project, 347 provenance fields, 93 retrieve list of projects, 347 retrieve list of scheme allele identifiers, 344 F retrieve list of submissions, 348 filters, 226 retrieve list of supporting files uploaded for submission. 351 G retrieve locus record, 335 retrieve project information, 347 Genome Comparator, 266 genome filtering, 81 retrieve scheme allele designation records, 344 in silico hybridization, 84 retrieve scheme information, 338 in silico PCR, 81 retrieve specific allelic profile record, 340 groups retrieve submission correspondence, 351 scheme, 66 retrieve submission record, 349 retrieve user information, 346 Η upload submission supporting file, 351 auto allele definer, 177 hosts stop, 179 mapping, 39 autotagger, 175

#### I

identify allele sequence, 195 sequence type, 197 in silico hybridization, 84 in silico PCR, 81 isolate records, 187 isolate record options, 237

#### L

list query alleles, 203 loci, 3 modifying display, 242 locus adding, 45, 48, 49, 55, 79 copying existing record, 53 extended attributes, 57 Locus Explorer, 245 codon usage, 249 polymorphic sites, 245 translated sequences, 250 locus positions setting, 85

#### Μ

main results table options, 236 mapping hosts, 39 materialized views, 40 MLST, 3 MLST scheme adding, 78 mod\_perl, 39 modifying display loci, 242 schemes, 242

## 0

offline curation auto allele definer, 177 autotagger, 175 options, 234 isolate record, 237 main results table, 236 provenance fields, 239 query, 240

#### Ρ

partitioning

sets, 41 passwords setting, 36 first user, 37 performance caching schemes, 39 materialized views, 40 mod\_perl, 39 permissions, 31 locus curation, 34 scheme curation, 34 plugins enabling, 37 polymorphic sites Locus Explorer, 245 polymorphisms, 288 profile records, 193 profiles, 4 projects, 146 provenance field breakdown, 251 provenance fields extended attributes, 93 options, 239

# Q

query allele designations, 219 options, 240 scheme profiles, 206 sequence tags, 224

# R

records allele definition, 191 isolate, 187 profile, 193 sequence bin, 193 sequence tag, 192 RESTful interface client authorization, 101 rule-based queries, 76

# S

scheme breakdown, 258 groups, 66 scheme profiles browse, 205 query, 206 schemes, 3 adding, 60

caching, 39 modifying display, 242 sequence bin breakdown, 264 records, 193 sequence similarity determining, 210 sequence tag records, 192 status, 293 sequence tags, 4 count, 222 query, 224 sequence type identify, 197 sets, 4 partitioning, 41 stop auto allele definer, 179 autotagger, 179

## Т

translated sequences Locus Explorer, 250 two-field breakdown, 255

## U

unique combinations, 286 updates disabling, 38 user groups, 31 user types, 31 users adding, 103