BIGSdb Documentation

Release 1.30.0

Keith Jolley

Jan 18, 2021

Contents

1	Conc	epts and terms	3
	1.1	BIGSdb	3
	1.2	Loci	3
	1.3	Alleles	3
	1.4	Schemes	4
	1.5	Profiles	4
	1.6	Classification groups	4
	1.7	Sequence tags	4
	1.8	Sets	5
2	BIGS	Sdb dependencies	7
	2.1	Required packages	7
3	Incto	llation and configuration of BIGSdb	9
5	3.1		9
	3.2		10
	3.3		11
	3.4		12
	3.5		12
	3.6		13
	3.7	81	14
	3.8		15
	3.9		15
	3.10		16
	3.11		16
			16
4	Data	hose setur	21
4	Data	1	21 21
	4.1 4.2		21 22
	4.2 4.3	1 0	22 23
	4.5 4.4	0	25 36
	4.4 4.5		30 37
		8 8	37 37
	4.6 4.7		<i>31</i> 39
	4.7 4.8		
			41 42
	4.9	User authentication	42

	4.10	Setting up the admin user
		Configuring access to remote contigs 44
5	Adm	inistrator's guide 47
	5.1	Types of user
	5.2	User groups
	5.3	Curator permissions
	5.4	Locus and scheme permissions (sequence definition database)
	5.5	Controlling access
	5.6	Setting user passwords
	5.7	Setting the first user password
	5.8	Enabling plugins
	5.9	Temporarily disabling database updates
	5.10	Host mapping
	5.11	Improving performance
	5.12	Dataset partitioning
	5.13	Setting a site-wide users database
	5.14	Adding new loci
	5.15	Defining locus extended attributes
	5.16	Defining schemes
	5.17	Organizing schemes into hierarchical groups
	5.18	Setting up client databases
	5.19	Workflow for setting up a MLST scheme
	5.20	Automated assignment of scheme profiles
	5.21	Scheme profile clustering - setting up classification schemes
	5.22	Defining new loci based on annotated reference genome
	5.23	Genome filtering
	5.24	Setting locus genome positions
	5.25	Defining composite fields
	5.26	Extended provenance attributes (lookup tables)
	5.20	Sequence bin attributes
		1
	5.28	
	5.29	Exporting table configurations
	5.30	Authorizing third-party client software to access authenticated resources 108
	5.31	BLAST caches
	5.32	Config-specific file downloads
_	C	
6		tor's guide 111
	6.1	Adding new sender details
	6.2	Adding new allele sequence definitions 112
	6.3	Updating and deleting allele sequence definitions
	6.4	Retiring allele identifiers
	6.5	Un-retiring allele identifiers
	6.6	Updating locus descriptions
	6.7	Adding new scheme profile definitions
	6.8	Updating and deleting scheme profile definitions
	6.9	Retiring scheme profile identifiers
	6.10	Un-retiring scheme profile identifiers
	6.11	Adding isolate records
	6.12	Updating and deleting single isolate records
	6.13	Batch updating multiple isolate records
	6.14	Deleting multiple isolate records
	6.15	Retiring isolate identifiers

		Un-retiring isolate identifiers	
	6.17	Setting alternative names for isolates (aliases)	
		Linking isolate records to publications	
		Uploading sequence contigs linked to an isolate record	
	6.20	Batch uploading sequence contigs linked to multiple isolate records	
		Linking remote contigs to isolate records	
	6.22	Automated web-based sequence tagging	
		Projects	
	6.24	Isolate record versioning	164
7	Cura	ting data submitted via the automated submission system	169
	7.1	Alleles	170
	7.2	Profiles	178
	7.3	Isolates	184
0	Ofd:	no curretion toole	189
8			
	8.1	Automated offline sequence tagging	
	8.2	Defining exemplar alleles	
	8.3	Automated offline allele definition	
	8.4	Cleanly interrupting offline curation	
	8.5	Uploading contigs from the command line	195
9	Defin	ition downloads	197
	9.1	Allele sequence definitions	197
	9.2	Scheme profile definitions	
	_		
10			205
		Isolate records	
		Allele definition records	
	10.3	Sequence tag records	
		Profile records	
	10.5	Sequence bin records	213
11	Ouer	ying data	215
		Querying sequences to determine allele identity	
		Querying multiple sequences to identify allele identities	
		Searching for specific allele definitions	
		Browsing scheme profile definitions	
		Querying scheme profile definitions	
		Identifying allelic profile definitions	
	11.7	Batch profile queries	
	11.8		237
	11.9		241
			242
			$\frac{2}{260}$
			263
			266 266
12			279
	12.1		280
	12.2		281
	12.3		282
	12.4		283
	12.5	Deleting a user project	284

13	Priva	te records 2	287
	13.1		287
		Modifying private records	
	13.3		290
	15.5		
14	Data	analysis plugins 2	.91
			291
			297
			301
			304
			315
			321
		Γ	330
			334
	14.9	Interactive Tree of Life (iTOL)	338
	14.10	Locus explorer	342
	14.11	Microreact	347
	14.12	PhyloViz	351
	14.13	Polymorphisms	353
			356
			360
			363
			,05 866
	14.17		000
15	Data	export plugins 3	371
15		Contig export	
		Isolate record export 3 Des Class sector 3	
		Profile export	
	15.4	Sequence export	579
16	Subr	nitting data using the submission system 3	383
10			384
			884
			393
		Isolate submission	
			-01
	16.6	Removing submissions from your notification list	02
17			103
	17.1		03
	17.2		03
	17.3	Resources	04
	17.4	Authentication	33
18	Frequ	uently asked questions (FAQs) 4	39
	18.1	General	39
	18.2	Installation	40
	18.3	Administration	40
19	Appe		43
	19.1	Query operators	43
	19.2	Sequence tag flags	44
	19.3	Allele sequence flags	45
20	Datal	base schema	47

Index

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 Illumina or Oxford Nanopore).

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 https://bigsdb.readthedocs.org/.

CHAPTER 1

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 Classification groups

Classification groups are a way to cluster scheme profiles using a specified threshold of pairwise allelic mismatches. Currently, single-linkage clustering is supported whereby each member of a group must have no more than the specified number of allelic differences with at least one other member of the group.

1.7 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

1.8 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'.

CHAPTER 2

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 9.5+
- 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.

• Archive::Zip - Used to upload to iTOL.

- Bio::Biblio This used to be part of BioPerl but will need to be installed separately if using BioPerl 1.6.920 or later.
- CGI (version 4.04+) Common Gateway Interface requests and responses (used to be a core module but recently removed).
- Config::Tiny Configuration file handling.
- Crypt::Eksblowfish::Bcrypt Used for password hashing.
- Data::Random
- Data::UUID Globally unique identifer handling for preference storage.
- DBD::Pg PostgreSQL database driver for DBI.
- DBI Database independent interface module used to interact with databases.
- Email::MIME Used to format E-mail messages.
- Email::Sender Used to send E-mail messages by submission system.
- Email::Valid Used to validate E-mails sent by job manager.
- Excel::Writer::XLSX Used to export data in Excel format.
- Exception::Class Exception handing.
- File::Type Used to determine what type of file has been uploaded.
- IO::String
- JSON Used to manipulate JSON data.
- List::MoreUtils (version 0.28+).
- Log::Dispatch::File Object for logging to file.
- Log::Log4perl Configurable status and error logging.
- LWP::UserAgent Used to upload via API
- Net::Oauth Required for REST authentication (this needs to be installed even if you are not using REST).
- Parallel::ForkManager Required for multi-threading tools and plugins.
- Time::Duration Used by Job Viewer to display elapsed time in rounded units.
- Try::Tiny
- XML::Parser::PerlSAX part of libxml-perl Used to parse XML configuration files.

2.1.3 Optional packages

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

- 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.

CHAPTER 3

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 javascript directory to the root directory of your website, i.e. accessible from http://your_website/ javascript/.
- 5. Copy the css directory to root directory of your website, i.e. accessible from http://your_website/css/.
- Copy the webfonts directory to the root directory of your website, i.e. accessible from http://your_website/ webfonts/.
- 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/

3.2.1 pg_hba.conf

```
# Database administrative login by UNIX sockets
local all postgres ident map=mymap
# TYPE DATABASE USER CIDR-ADDRESS METHOD
# "local" is for Unix domain socket connections only
local all all ident map=mymap
```

(continues on next page)

(continued from previous page)

# IPv4	local	connections:				
host	all	all	127.0.0.1/32	md5		
# IPv6	5 local	connections:				
host	all	all	::1/128	md5		

3.2.2 pg_ident.conf

# MAPNAME	SYSTEM-USERNAME	PG-USERNAME
mymap mymap mymap	postgres webmaster www-data bizadb	postgres postgres apache bizadh
mymap mymap	bigsdb bigsdb	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 Setting global connection parameters

Global database connection parameters can be entered in /etc/bigsdb/db.conf. This allows you to set default values for the host, port, user and password. Default values are as follows:

- · dbhost: localhost
- dbport: 5432
- dbuser: apache
- dbpassword: remote

These can all be over-ridden in individual *database configuration config.xml files* using the terms host, port, user, and password.

3.4 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.

Breadcrumb navigation links can be configured with a file called breadcrumbs.conf, placed either in the database configuration directory, the root directory of the website, or in /etc/bigsdb/conf. The file describes links that are higher in the hierarchy than the database index page. The file consists of lines that contain link text separated by a pipe symbol (I) followed by a URL for that link, e.g.

```
Home|/
Organisms|/databases/
```

3.5 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.
 - Copy bigsjobs.pl to /usr/local/bin

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.pl

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

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
```

3.6 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.

1. 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.

3.7 Setting up a site-wide user database

A site-wide user database allows users to register themselves for accounts and associate these with specific databases. It means that a single set of log-in credentials can be used across databases, rather than each database maintaining its own.

Users can access/update their account details by calling the bigsdb.pl script without any additional attributes, e.g. http://website/cgi-bin/bigsdb.pl.

Site admins can access administration features by calling the bigscurate.pl script without any additional attributes.

1. Create a user database, e.g. pubmlst_bigsdb_users:

```
createdb pubmlst_bigsdb_users
psql -f users.sql pubmlst_bigsdb_users
```

Set up sync_user_dbase_users.pl to run every hour as a CRON JOB, e.g. in /etc/crontab, add the following to run this at 5 minutes past each hour

```
05 * * * * bigsdb /usr/local/bin/sync_user_dbase_users.pl --user_database_

→pubmlst_bigsdb_users
```

Add the user database details to each database that you want to allow to use it.

You need to add the users database details to each client database that will use it.

2. If you want to allow users to register themselves you need to modify bigsdb.conf.

You can define multiple user databases (as a comma-separated list) but usually you would have just one. Define this using the site_user_dbs attribute. Use a short domain (site) name separated by a pipe (l) and the name of the database, e.g. add the following to /etc/bigsdb.conf:

site_user_dbs=PubMLST|pubmlst_bigsdb_users

Make sure default database connection parameters are set in /etc/bigsdb/db.conf.

3. Set up site admin user in new user database. This has to be done manually - other users will either be able to register themselves or be created by curators from other databases.:

```
psql pubmlst_bigsdb_users
INSERT INTO USERS (user_name,surname,first_name,email,affiliation,
    date_entered,datestamp,status) VALUES ('kjolley','Jolley','Keith',
    'keith.jolley@zoo.ox.ac.uk','University of Oxford, UK','now','now',
    'validated');
```

Set the password for this user using the add_user.pl script (change XXXXXXXX to the password value):

add_user.pl -a -d pubmlst_bigsdb_users -n kjolley -p XXXXXXXX

Add specific permissions that this admin user can have by directly adding the following terms to the permissions table:

- set_site_user_passwords:
 - Allow admin to set user passwords.
- import_dbase_configs:
 - Allow admin to define which database configurations are made available for registration.
- merge_users

- Allow admin to merge user accounts.
- modify_users
 - Allow admin to edit user details.

e.g.

```
psql pubmlst_bigsdb_users
INSERT INTO permissions (user_name,permission,curator,datestamp) VALUES
  ('kjolley','import_dbase_configs','kjolley','now');
```

- 4. Specific *permissions can be set for curators* in individual databases:
 - import_site_users
 - This allows the curator to import site users in to the database.
 - modify_site_users
 - You may not wish to do this! It allows the curator of any database with this permission to change the details of a user that may be used on other databases on the site.
- 5. HTML header files can be defined for use when bigsdb.pl or bigscurate.pl are called without a database configuration, such as when a user is registering or modifying their user details. These files, site_header.html, site_footer.html, site_curate_header.html and site_curate_footer.html should be placed in the root directory of the web site.

3.8 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
```

3.9 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).

3.10 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 {
 weekly
 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
```

3.11 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 -> 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.

3.12 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 Starman -E deployment
end script
```

The service will then start automatically on boot or can be manually started by calling:

sudo service bigsdb-rest start

For systems using systemd, create a file in /etc/systemd/system called bigsdb-rest.service with the following contents (again, modify file paths as appropriate):

```
[Unit]
Description=BIGSdb REST interface
After=network.target
[Service]
User=bigsdb
ExecStart=/usr/bin/plackup -a /var/rest/bigsrest.pl -s Starman -E deployment
Restart=always
```

(continues on next page)

(continued from previous page)

```
[Install]
WantedBy=multi-user.target
```

To start the service automatically on boot you need to enable it:

sudo systemctl enable bigsdb-rest.service

It can also be manually started by calling:

```
sudo systemctl start bigsdb-rest.service
```

3.12.1 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>
```

You should also set 'rest_behind_proxy=1' in bigsdb.conf.

CHAPTER 4

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.

The xml directory of the software archive contains example XML files for sequence definition and isolate databases (rename these to config.xml). The isolates_config.xml file contains the minimum required isolate table fields and matches the isolate table that will be generated using the isolatedb.sql SQL script.

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.
- profile_submit.html HTML markup for text that is inserted in to the submission interface prior to profile submission finalization. This can be used to add specific instructions such as the requirement to make an isolate submission.
- allele_submit.html HTML markup for text that is inserted in to the submission interface prior to allele submission finalization. This can be used to add specific instructions such as the requirement to attach Sanger trace files.

- isolate_submit.html HTML markup for text that is inserted in to the submission interface prior to isolate submission finalization. This can be used to add specific instructions such as the request to also make a new profile submission if the isolate has a new profile.
- profile_curate.html HTML markup for text that is inserted on submission curation page if profile submissions
 are pending. This can be used to add specific information to curators.
- allele_curate.html HTML markup for text that is inserted on submission curation page if allele submissions
 are pending. This can be used to add specific information to curators.
- isolate_curate.html HTML markup for text that is inserted on submission curation page if isolate submissions are pending. This can be used to add specific information to curators.
- registration_success.txt Text file containing message content to be used in an automated E-mail when granting access to a user who has requested access to the database using the site-wide account system (where auto-registration is not enabled).

The header and footer files can alternatively be placed in the root directory of the web site, or in /etc/bigsdb, for site-wide use. If files exist in multiple locations, they are used in the following order of preference: database config directory > web root directory > /etc/bigsdb.

There are four additional files, site_header.html, site_footer.html, curate_site_header.html and curate_site_footer.html which are used when either bigsdb.pl or bigscurate.pl are called without a database configuration. These should be placed in the root directory of the web site or in /etc/bigsdb.

You can also add HTML meta attributes (such as a favicon) by including a file called meta.html in the database configuration directory. For example to set a favicon this file can contain something like the following:

```
<link rel="shortcut icon" href="/favicon.ico" type="image/ico" />
```

These attributes will appear in the <head> section of the HTML page.

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 elements must match the physical structure of the database isolate table. Required attributes are in **bold**:

<db>

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

<system>

Any value set here can be overridden in a system.overrides file.

- 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 (see also 'formatted_description').
- 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.
- BLAST
 - Enable Blast plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the Blast plugin can be disabled by setting this attribute to 'no'.
- BURST
 - Enable BURST plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the BURST plugin can be disabled by setting this attribute to 'no'.
- cache_schemes
 - Enable automatic refreshing of scheme field caches when batch adding new isolates: either 'yes' or 'no', default 'no'.
 - See scheme caching.
- CodonUsage
 - Enable Codon Usage plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the Codon Usage plugin can be disabled by setting this attribute to 'no'.
- 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'.
- ContigExport
 - Enable contig export plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the contig export plugin can be disabled by setting this attribute to 'no'.
- curate_config
 - The database configuration that should be used for curation if different from the current configuration. This is used when the submission system is being used so that curation links in the 'Manage submissions' pages for curators load the correct database configuration.
- curate_link

- URL to curator's interface, which can be relative or absolute. This will be used to create a link in the public interface dropdown menu.
- 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.
- curators_only
 - Set to 'yes' to prevent ordinary authenticated users having access to database configuration. This is only effective if read_access is set to 'authenticated_users'. This may be useful if you have different configurations for curation and querying with some data hidden in the configuration used by standard users. Default 'no'.
- daily_pending_submissions
 - Overrides the daily limit on pending submissions that a user can submit via the web submission system. Default: '15'.
- daily_rest_submissions_limit
 - Overrides the limit on number of submissions that can be made to the database via the RESTful interface. This is useful to prevent flooding of the submission system by aberrant scripts. Default: '100'.
- 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_private_records
 - The default number of private isolate records that a user can upload. The user account must have a status of either 'submitter', 'curator', or 'admin'. This value is used to set the private_quota field when creating a new user record (which can be overridden for individual users). Changing it will not affect the quotas of existing users. Default: '0'.
- 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').
- delete_retire_only
 - Set to 'yes' to retire the id of any isolate that is deleted. This prevents re-use of ids. This setting will override the global setting in bigsdb.conf.
- disable_updates

- Set to 'yes' to prevent updates. This is useful when moving databases or temporarily running on a backup server.
- disable_update_message
 - Message shown when updates are disabled.
- eav_fields
 - Name to call sparsely-populated fields. Default: 'secondary metadata'.
- eav_field_icon
 - Icon class from FontAwesome to use on isolate info page for sparsely- populated fields. Default 'fas fa-microscope'.
- eav_groups
 - Comma-separated list of category names that sparsely-populated fields can be grouped in to. If this value is set, a category drop-down list will appear when adding or updating sparsely-populated fields. You can add an icon to appear by following the name with a pipe symbol (I) and an icon class from the FontAwesome library, e.g. 'Vaccine reactivitylfas fa-syringe,Risk factors/fas fa-smoking'.
- export_limit
 - Overrides the default allowed number of data points (isolates x columns) to export. Default: '25000000'.
- fast_scan
 - Sets whether fast mode scanning is enabled via the web interface. This will scan all loci together, using exemplar sequences. In cases where multiple loci are being scanned this should be significantly faster than the standard locus-by-locus scan, but it will take longer for the first results to appear. *Allele exemplars* should be defined if you enable this option. Set to 'yes' to enable. Default: 'no'.
- 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.
- formatted_description
 - Markdown formatted description of database. If set, this will be used throughout the HTML interface wherever formatting can be applied (main body of text) and overrides the value set in 'db_description'. Currently only supports **italics** and ****bold****.
- genepresence_record_limit
 - Overrides the record number limit (isolates x loci) for the Gene Presence plugin. Default: 500000 (this can also be set globally in bigsdb.conf).
- genepresence_taxa_limit
 - Overrides the isolate limit for the Gene Presence plugin. Default: 10000 (this can also be set globally in bigsdb.conf).
- GenomeComparator
 - Enable Genome Comparator plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the Genome Comparator plugin can be disabled by setting this attribute to 'no'.
- genome_comparator_limit
 - Overrides the isolate number limit for the Genome Comparator plugin. Default: 1000 (this can also be set globally in bigsdb.conf).

- genome_comparator_max_ref_loci
 - Overrides the limit on number of loci allowed in a reference genome. Default: 10000.
- genome_comparator_threads
 - The number of threads to use for data gathering (BLAST, database queries) to populate data structure for Genome Comparator analysis. You should not set this to less than 2 as this will prevent job cancelling due to the way isolates are queued. Default: '2'.
- 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'.
- itol_record_limit
 - Overrides the maximum number of records that can be included in an ITOL job. Default: 2000 (this can also be set globally in bigsdb.conf).
- itol_seq_limit
 - Overrides the maximum number of sequences (records x loci) that can be included in an ITOL job.
 Default: 100,000 (this can also be set globally in bigsdb.conf).
- 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.
- Microreact
 - Enable Microreact plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the Microreact plugin can be disabled by setting this attribute to 'no'. Note that for the plugin to be active, a country field containing a defined list of allowed values and an integer year field must be defined in the isolates table.
- microreact_record_limit

- Overrides the maximum number of records that can be included in a Microreact job. Default: 2000 (this can also be set globally in bigsdb.conf).
- microreact_seq_limit
 - Overrides the maximum number of sequences (records x loci) that can be included in an Microreact job.
 Default: 100,000 (this can also be set globally in bigsdb.conf).
- min_genome_size
 - Size in bp that is the minimum size of the sequence bin considered to represent a whole genome. This is used in the REST interface to differentiate records with genomes. You can also pass a 'genomes=1' attribute to the an isolate query form and this will populate the appropriate search to return genome records.
- new_version
 - Set to 'no' to prevent copying field value when creating a new version of the isolate record.
- 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'.
- pcr_limit
 - Overrides the isolate number limit for the in silico PCR plugin. Default: '10000'.
- PhyloViz
 - Enable third party PhyloViz plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the PhyloViz plugin can be disabled by setting this attribute to 'no'.
- 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'.
- public_login
 - Optionally allow users to log in to a public database this is useful as any jobs will be associated with the user and their preferences will also be linked to the account. Set to 'no' to disable. 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'.
- related_databases
 - Semi-colon separated list of links to related BIGSdb databases on the system. This should be in the form of database configuration name followed by a 'l' and the description, e.g. 'pubmlst_neisseria_seqdeflSequence and profile definitions'. This is used to populate the dropdown menu.
- remote_contigs
 - Optionally allow the use of remote contigs. These are stored in a remote BIGSdb database, accessible via the RESTful API. Set to 'yes' to enable.
- rest_kiosk
 - If 'kiosk' attribute is set, then the REST interface will be disabled for the configuration unless a value is set here. The only supported value currently is 'sequenceQuery' which will enable API routes for querying sequences.
- rMLSTSpecies
 - Enable rMLST Species identifier plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the plugin can be disabled by setting this attribute to 'no'. Note that for the plugin to be active, a country field containing a defined list of allowed values and an integer year field must be defined in the isolates table.
- script_path_includes
 - Partial path of the bigsdb.pl script used to access the database. See *user authentication*.
- · SeqbinBreakdown
 - Enable Sequence bin breakdown plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the plugin can be disabled by setting this attribute to 'no'. Note that for the plugin to be active, a country field containing a defined list of allowed values and an integer year field must be defined in the isolates table.
- 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.
- submissions_deleted_days

- Overrides the default number of days before closed submissions are deleted from the system. Default: '90'.
- TagStatus
 - Enable Tag status plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the plugin can be disabled by setting this attribute to 'no'. Note that for the plugin to be active, a country field containing a defined list of allowed values and an integer year field must be defined in the isolates table.
- tblastx_tagging
 - Sets whether tagging can be performed using TBLASTX: either 'yes' or 'no', default 'no'.
- total_pending_submissions
 - Overrides the total limit on pending submissions that a user can submit via the web submission system. Default: '20'.
- user
 - Username for access to isolates database, default 'apache'.
- user_job_quota
 - Integer with number of offline jobs that can be queued or currently running for this database by any specific user - this parameter is only effective if users have to log in.
- user_projects
 - Sets whether authenticated users can create their own projects in order to group isolates: either 'yes' or 'no', default 'no'.
- 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, or to restrict loci or schemes to a subset of isolate data.
- webroot
 - URL of web root, which can be relative or absolute. This is used to provide a hyperlinked item in the dropdown menu. Default '/'.
- webroot_label
 - Label text for the breadcrumb link defined by the webroot value. This can be formatted using Markdown. Currently only supports **italics** and ***bold***.

<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>
```

(continues on next page)

(continued from previous page)

```
<option>pandemic</option>
</optlist>
</field>
```

• type

- Data type: int, text, float, bool, or date.
- allow_submissions
 - Show in submission template and allow data to be submitted even if field is set as 'curate_only'. This has no effect on fields that do not have the 'curate_only' attribute as these fields are included in submissions by default. This attribute will be overridden if the field has the 'no_submissions' attribute set.
- comments
 - Comments about the field. These will be displayed in the field description plugin and as tooltips within the curation interface.
- curate_only
 - Set to 'yes' to hide field unless logged-in user is a curator or admin. Set the 'allow_submissions' attribute to still include the field in the submission template so that it can be included in submissions of new records by standard users.
- default
 - Default value. This will be entered automatically in the web form but can be overridden.
- 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.
- log_delete
 - Sets if the field value will be recorded in the log table if the isolate is deleted. Set to 'yes' or 'no', default is 'no'. The id and isolate name are always recorded if deletion is logged.
- 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 and date types. Special values such as CURRENT_YEAR and CUR-RENT_DATE can be used.
- min
 - Minimum value for integer and date types.
- multiple
 - Sets if field allows multiple values to be set for it, 'yes' or 'no', default 'no'. If set to 'yes', then the underlying field in the database must be an ARRAY type, e.g. text[].
- no_curate

- Setting this will hide the field in the curator interface and prevent it from being manually modified. This is useful for fields that are populated by automated scripts or database triggers. Can be 'yes' or 'no', default 'no'.
- no_submissions
 - Setting this will hide the field in the submission template. The field is still available if it is added back to the template manually.
- 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', 'no', or 'expected', default 'yes'. If set to 'expected', the value cannot be left empty when batch adding an isolate record or using the submission system, but a null value can be explicitly set using the value 'null'. The use of this is to encourage submitters to include a value for this field if it is available, while still allowing empty values if it is not.
- 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_DATE: current date in yyyy-mm-dd format
- CURRENT_YEAR: the 4 digit value of the current year

4.3.2 Sequence definition database XML attributes

Required attributes are in **bold**.

```
<db>
```

Top level element. Contains child element: system.

<system>

Any value set here can be overridden in a system.overrides file.

- 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'.
- BURST
 - Enable BURST plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all_plugins attribute is set to 'yes'. If the all_plugins attribute is set to 'yes', the BURST plugin can be disabled by setting this attribute to 'no'.
- curate_config
 - The database configuration that should be used for curation if different from the current configuration. This is used when the submission system is being used so that curation links in the 'Manage submissions' pages for curators load the correct database configuration.
- 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.
- curators_only
 - Set to 'yes' to prevent ordinary authenticated users having access to database configuration. This is only
 effective if read_access is set to 'authenticated_users'. This may be useful if you have different configurations for curation and querying with some data hidden in the configuration used by standard users. Default
 'no'.
- daily_pending_submissions
 - Overrides the daily limit on pending submissions that a user can submit via the web submission system. Default: '15'.
- daily_rest_submissions_limit
 - Overrides the limit on number of submissions that can be made to the database via the RESTful interface. This is useful to prevent flooding of the submission system by aberrant scripts. Default: '100'.
- delete_retire_only
 - Set to 'yes' to retire the id of any allele or profile that is deleted. This prevents re-use of ids. This setting will override the global setting in bigsdb.conf.

- 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'.
- exemplars
 - Use exemplar sequences in the BLAST caches used for the sequence query pages. This is useful on larger databases as it speeds up the query significantly. *Exemplar alleles MUST* be defined otherwise sequence queries will fail. 'yes' or 'no', default 'no'.
- formatted_description
 - Markdown formatted description of database. If set, this will be used throughout the HTML interface wherever formatting can be applied (main body of text) and overrides the value set in 'db_description'. Currently only supports **italics** and ****bold****.
- isolate_database
 - The config name of the isolate database. This is used to provide a link to isolate submissions. You also need to set isolate_submissions="yes".
- isolate_submissions
 - Set to yes to provide a link to isolate submissions. The isolate_database attribute also needs to be set. 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.
- kiosk
 - Set to a page name to restrict configuration to always start on this page, rather than an index page. This faciliates running in a cut-down *kiosk mode* that doesn't allow access to all features. *Currently only 'sequenceQuery' is supported*.
- kiosk_allowed_pages
 - Comma-separated list of pages that the configuration is allowed to show, apart from the page set in the 'kiosk' attribute. Example for a sequence query configuration would be 'sequenceTranslate' to allow access to the translated sequence page following a query.
- kiosk_help
 - URL to context-sensitive help page.
- kiosk_locus
 - Restrict sequence query to a specific locus or scheme. Use either the locus primary name or 'SCHEME_X' where X is the scheme number.
- kiosk_no_genbank
 - Set to "yes" to hide the Genbank accesssion form element in kiosk mode.
- kiosk_no_upload

- Set to "yes" to hide the sequence file upload in kiosk mode.
- kiosk_simple
 - Remove most explanatory text from kiosk page.
- kiosk_text
 - Alternative text to show on kiosk page.
- kiosk_title
 - Title text to use when running in kiosk mode.
- 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.
- public_login
 - Optionally allow users to log in to a public database this is useful as any jobs will be associated with the user and their preferences will also be linked to the account. Set to 'no' to disable. 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'.
- related_databases
 - Semi-colon separated list of links to related BIGSdb databases on the system. This should be in the form of database configuration name followed by a 'l' and the description, e.g. 'pubmlst_neisseria_isolates/Isolates'. This is used to populate the dropdown menu.
- 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.
- submissions_deleted_days
 - Overrides the default number of days before closed submissions are deleted from the system. Default: '90'.
- total_pending_submissions
 - Overrides the total limit on pending submissions that a user can submit via the web submission system. Default: '20'.
- user_job_quota
 - Integer with number of offline jobs that can be queued or currently running for this database by any specific user this parameter is only effective if users have to log in.
- webroot
 - URL of web root, which can be relative or absolute. This is used to provide a hyperlinked item in the dropdown menu. Default '/'.
- webroot_label
 - Label text for the breadcrumb link defined by the webroot value. This can be formatted using Markdown. Currently only supports **italics** and ***bold***.

4.4 Over-riding global defaults set in bigsdb.conf

Certain values set in bigsdb.conf can be over-ridden by corresponding values set in a database-specific config.xml file. These can be set within the system tag like other attributes:

- query_script
 - Relative web path to bigsdb script.
- curate_script
 - Relative web path to curation script.
- prefs_db
 - The name of the preferences database.
- auth_db
 - The name of the authentication database.
- tmp_dir
 - Path to the web-accessible temporary directory.
- secure_tmp_dir
 - Path to the web-inaccessible (secure) temporary directory.
- ref_db
 - The name of the references database.

4.5 Over-riding values set in config.xml

Any attribute used in the system tag of the database config.xml file can be over-ridden using a file called **system.overrides**, placed in the same directory as config.xml. This is very useful as it allows you to set up multiple configs for a database, with the config.xml files symlinked so that any changes to one will be seen in each database configuration. An example of why you may wish to do this would be if you create separate public and private views of the isolate table that filters on some attribute. The system.overrides file uses key value pairs separated by = with the values quoted, e.g.

```
view="private"
read_access="authenticated_users"
description="Private view of database"
```

It is also possible to override the required, maindisplay or curate_only attributes of a particular field using a file called **field.overrides**. The field.overrides file uses the format 'field:attribute="value" on each line, e.g.

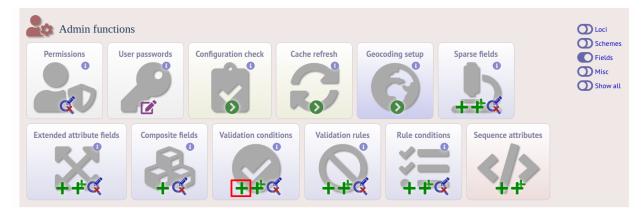
date_received:required="yes"

4.6 Setting field validation rules

Sometimes it may be necessary to restrict the allowed values in one isolate field depending on the values submitted for another field. It is possible to do this using field validation rules. These combine one or more conditions which all have to match for validation to fail and an isolate record upload to be rejected.

An example of this may be if you have an age_year and an age_month field but you only want age_month to be populated if the subject is less than one year old. You can do this as follows.

As an admin, on the curator interface, click the 'Field' toggle to show the validation table links. Then click 'Add' on the 'Validation conditions' setting:



Add the following conditions separately:

- $age_year > 0$
- age_month NOT null

PubMLST		
Home > Organisms > Organism > Neisseria isolates > Add new validation condition		
Add new validation condition		
Please fill in the fields below - required fields are marked with an exclamation mark ().	Action RESET SUBMIT	ľ

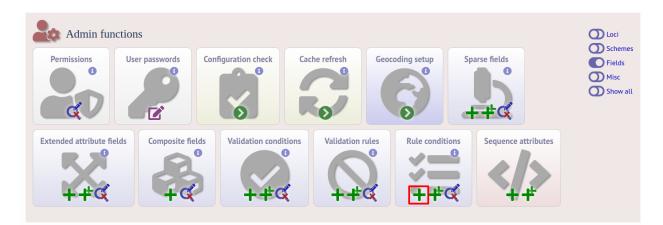
Now add a new 'Validation rule', by clicking 'Add' on the 'Validation rules' setting:

Admin functi	ons						O Loci
Permissions	User passwords	Configuration check	Cache refresh	Geocoding	setup S	parse fields	 Schemes Fields Misc Show all
Extended attribute fields	Composite field			5	ule conditions	Sequence attributes	

Here you just enter the message that will be returned when the validation fails:

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Add new validation rule	
Add new validation rule	8
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record id: 1 message: lage_mth should only be set when age_yr < 1. curator:! Keith Jolley (keith) datestamp:: 2020-07-18	ľ

Finally add the conditions to the rule by clicking 'Add' on the 'Rule conditions' setting:



Select the rule message and the condition from the dropdown boxes:

PubMLST		
Home > Organisms > Organism > Neisseria isolates > Add new rule condition		
Add new rule condition		2
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record rule id: 1) age mth should only be set when age yr < 1.	Action	
condition id: 1) age_yr>0 v curator: Keth Jolley (keth) datestamp: 2020-07-18	RESET SUBMIT	

Make sure you do this for each of the conditions that have to match.

Validation checks are performed when adding or updating an isolate record, or when a user submits via the automated submission interface. Currently these checks are not enforced when doing a batch update.

4.6.1 Special condition values

Use the value null to indicate that the field is empty, e.g.

• age_month NOT null

Use a field name in square brackets to compare the value in that field, e.g. suppose you have two date fields, 'date_sampled' and 'date_received', and you want to ensure that 'date_received' is not before 'date_sampled'. You can do this with the following condition:

date_received < [date_sampled]

The two fields have to be of the same data type in order to be compared (you cannot compare a text field to an integer field for example).

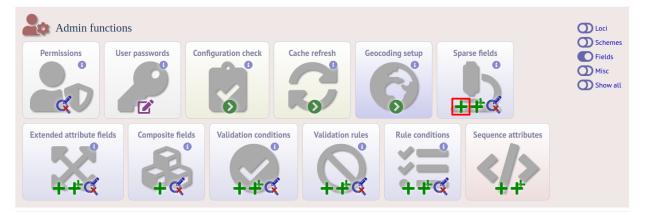
4.7 Sparsely-populated fields

Commonly used isolate fields should be described in the config.xml file and included as columns within the isolates table. Sometimes, however, you may have a need to record information that is only likely to be found in a minority of records. This can be done more efficiently with the use of sparsely-populated fields. These are stored differently in

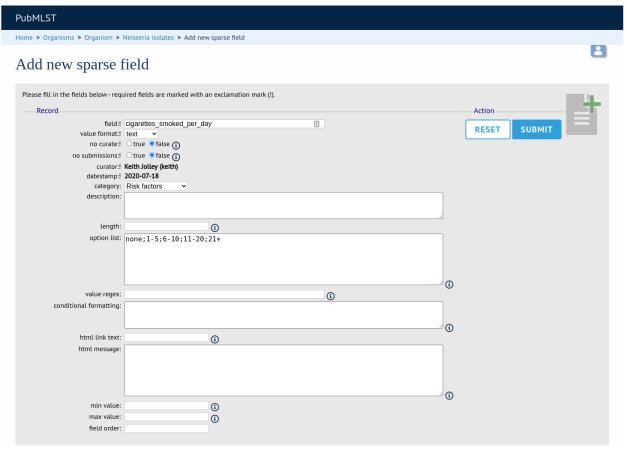
the database (using an entity-attribute-value [EAV] model) but can still be searched and exported in a similar way to normal fields. There is no limit to the number of such fields that can be defined.

The default name for these fields is 'secondary metadata' and this is how they will be grouped in the interface. You can change this by setting the 'eav_fields' attribute in the *system tag of config.xml*. It is also possible to group these fields in to categories - these can be defined with a comma-separated list in the 'eav_groups' attribute in the *system tag of config.xml*.

You will need to be an admin to define sparely-populated fields. Make sure that the 'Fields' toggle is selected on the curators' page. Click the add (+) button on the 'Sparse fields' function.



Fill in the form and click 'Submit'.



Field options are:

• field

- name of field
- value_format
 - date type either integer, float, text, date or boolean.
- no_curate
 - Set to true to prevent user updates of fieldThis setting could be used if the value is calculated by an external script rather than entered by a curator.
- no_submissions
 - Set to true to prevent the field being listed in the submissions template.
- description
 - Tooltip text that will appear on curator forms.
- length
 - Restrict allowed length of value.
- option_list
 - Semi-colon separated list of allowed values.
- value_regex
 - Regular expression that can constrain allowed values.
- conditional_formatting
 - Semi-colon separated list of values each consisting of the value, followed by a pipe character (I) and HTML to display instead of the value. If you need to include a semi-colon within the HTML, use two semi-colons (;;) otherwise it will be treated as the list separator.'
- html_link_text
 - This defines the text that will appear on an information link that will trigger a slide-in message (if defined int the next field). Default is 'info'.
- html_message
 - This message will slide-in on the isolate information page when the field value is populated and the information link is clicked. Full HTML formatting is supported.
- min_value
 - Valid for number fields only.
- max_value
 - Valid for number fields only.
- field_order
 - Integer indicating the order that fields should be displayed. If this is not set they will appear alphabetically.

4.8 Kiosk mode

Kiosk mode allows you to run a cut-down interface that offers a single main functionality. Currently, only a sequence query page is supported. The interface is locked down so that only specified functionality is supported and data cannot be exported.

See the *kiosk_* attributes* in config.xml.

As an example, the following settings are used for the rMLST 'Identify species' tool at https://pubmlst.org/rmlst/. The database usually requires a user to log in, but this tool offers a restricted functionality without logging in.

```
kiosk="sequenceQuery"
kiosk_allowed_pages="sequenceTranslate"
kiosk_title="Identify species"
kiosk_locus="SCHEME_1"
kiosk_simple="yes"
kiosk_no_upload="no"
kiosk_no_genbank="no"
rest_kiosk="sequenceQuery"
```

When you go to this example kiosk page you see only the sequence query page and trying to access any other functionality is prevented.

The rest_kiosk attribute enables queries to also be performed using the *RESTful API* which will be similarly locked down.

BIGSdb		
Home > Organisms > Organism > Identify species		
Identify species		
Please paste in your sequence to query against the database.		
Enter query sequence (single or multiple contigs up to whole genome in size)	Alternatively upload FASTA file	or enter Genbank accession
	Select FASTA file: Choose file No file chosen	Action
		RESET SUBMIT

4.9 User authentication

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

4.9.1 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'.

4.9.2 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. Stored passwords are salted and hashed using bcrypt.

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

4.10 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).:

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).

4.11 Retrieving PubMed citations from NCBI

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 retrieve_pubmed_records.pl and can be found in the scripts/maintenance directory.

Simply run the script 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.

4.12 Configuring access to remote contigs

It is possible for isolate records to have contigs in an external BIGSdb database. These must be accessible via the *RESTful API*. The advantage of this is that it enables multiple isolate databases to use the same genome assemblies without having to duplicate the storage of those assemblies. If access to the external database requires authenticated access, OAuth settings can be set to enable contig retrieval.

To enable remote contigs, set the remote_contigs attribute in the *<system>* tag of config.xml, i.e.

```
remote_contigs = "yes"
```

4.12.1 Setting up authentication

A client key for the BIGSdb remote contig manager needs to be generated. This can be done using the *cre-ate_client_credentials.pl* script, e.g.

create_client_credentials.pl --a 'BIGSdb remote contig manager' --insert

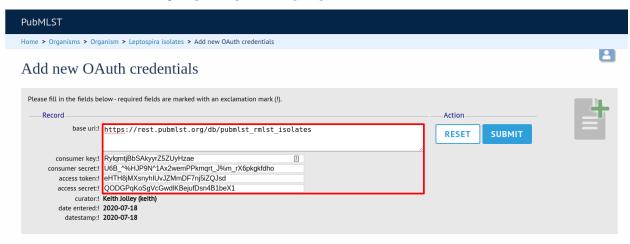
This will generate a client id (key) and a client secret and add them to the authentication database.

You will then need to obtain an access token and access secret using the client key and secret with the get_oauth_access_token.pl script. You will need to enter the API database URI (e.g. http://rest.pubmlst.org/db/pubmlst_rmlst_isolates) and the web database URL (e.g. https://pubmlst.org/bigsdb?db=pubmlst_rmlst_isolates). You will then be prompted to follow a link and log in to the database with your user credentials. A verification code will be generated. You need to enter this in to the script when prompted. An access token and secret will be returned to you.

From the curators' page, click the oauth credentials add link in the administrator settings. This function is normally hidden, so you may need to click the 'Misc' toggle to display it.



Populate the OAuth_credentials table with the client key/secret and access token/secret. You should also enter the root REST URI for the database (e.g. http://rest.pubmlst.org/db/pubmlst_rmlst_isolates).



4.12.2 Processing remote contigs

When remote contigs are first linked to a record, the sequences are downloaded in bulk (without their metadata). This allows the sequence lengths to be recorded as this is needed for various queries and outputs. The curator is then given an option to process the contigs, which involves downloading each contig individually to record metadata including the original designation and the sequence platform used. This may take a while so it may be preferable to perform this task offline. This can be done using the process_remote_contigs.pl script found in the scripts/automation directory. Options for using this script are shown below:

```
remote_contigs.pl --help
NAME
    process_remote_contigs.pl
   Download, check length and create checksum contigs stored as URIs
SYNOPSIS
   process_remote_contigs.pl --database NAME [options]
OPTIONS
--database NAME
   Database configuration name.
--exclude_isolates LIST
    Comma-separated list of isolate ids to ignore.
--exclude_projects LIST
   Comma-separated list of projects whose isolates will be excluded.
--help
   This help page.
--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).
--min ID
   Minimum isolate id.
--max ID
   Maximum isolate id.
--projects LIST
   Comma-separated list of project isolates to scan.
--quiet
    Only display errors.
```

CHAPTER 5

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.
- query_users allowed to query and view users registered to the database. This is automatically allowed if permission is set to modify users.
- modify_users allowed to add or modify user records. They can change the status of users, but can not revoke admin privileges 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_locus_descriptions allowed to modify the description text and external hyperlinks used. Even with this setting, only loci for which a user is explicitly set as a curator can be modified.
- 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).
- import_site_users allowed to import site users in to the database.
- modify_site_users allowed to modify site user details (you may not want to this! The user account can be used by multiple databases on the site and any changes to user details will be seen throughout the site).

Permissions can be set by clicking the Update/delete button in the Permissions box in the admin functions area of the curator's interface:



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

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Set curator permissions	
Set curator permissions	L Help 🗹
Select curator(s) Action de Korne, Joinna (ukorne) Debech, Nadia (ndebech) Deghmane, Ala-Eddin (deghmane) SELECT Demzuk, Walter (wdemzuk) Desmet, Stefanie (sdsmet6) Dialio, Kanry (kdialio) Dialio, Kanry (kdialio) Dillon, Jo-Anne (dillon) Exley, Rachel (rexley)	

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

> Organisms > Organism CUITATOT PET Select curator(s) ehovin, Ana (acehovin) han, Hannah (hchan) hen, Mingliang (cmibrigh hoi, Eunhwa (eunchoi)	missions	Acti	
Select curator(s) ehovin, Ana (acehovin) han, Hannah (hchan) hen, Mingliang (cmlbrigh		Acti	on
Select curator(s) ehovin, Ana (acehovin) han, Hannah (hchan) hen, Mingliang (cmlbrigh		Acti	on
Select curator(s) ehovin, Ana (acehovin) han, Hannah (hchan) hen, Mingliang (cmlbrigh		Acti	on
ehovin, Ana (acehovin) han, Hannah (hchan) hen, Mingliang (cmlbrigh			on ———
ehovin, Ana (acehovin) han, Hannah (hchan) hen, Mingliang (cmlbrigh			on
han, Hannah (hchan) hen, Mingliang (cmlbrigh			
han, Hannah (hchan) hen, Mingliang (cmlbrigh			
hen, Mingliang (cmlbrigh		SEL	.ECT
ani Europune (numeroni)			
hristodoulides, Myron (m	nchristodoul)		
lark, Stephen (SClark)			
laus, Heike (hclaus)			
lemence, Marianne (mcl	emence)	•	
All No	one		
	Curator		
	emence, Marianne	Diallo, Kanny	
query users			
modify users			
modify isolates modify projects			
modify sequences			
tag sequences			
designate alleles			
modify usergroups			
set user passwords			
modify loci			
modify schemes			
	0		
		0	
modify composites			
modify composites nodify field attributes			
modify composites nodify field attributes nodify value attributes			
modify composites nodify field attributes nodify value attributes modify sparse fields			
modify composites nodify field attributes nodify value attributes modify sparse fields modify probes			
modify composites nodify field attributes nodify value attributes modify sparse fields modify probes modify experiments			
modify composites nodify field attributes nodify value attributes modify sparse fields modify probes modify experiments delete all			
modify composites nodify field attributes nodify value attributes modify sparse fields modify probes modify probes delete all import site users			
modify composites nodify field attributes nodify value attributes modify sparse fields modify probes modify experiments delete all import site users modify site users			
modify composites nodify field attributes nodify value attributes modify sparse fields modify probes modify probes delete all import site users			

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':



Select the curator from the list:

PubMLST	
Home > Organisms > Organism > Neisseria typing > Batch update locus curator access records	
Batch update locus curator access records	2
Select user The user status must also be set to curator for permissions to work. users: Jolley, Keith (keith) Select	

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:



Available		Selected		
FifR Regulon NEIS0001 NEIS0004 NEIS0005 NEIS0005 NEIS0006 NEIS0006 NEIS0007 NEIS0008 NEIS0008 NEIS0009 NEIS0010 NEIS0010 NEIS0010 NEIS0012 NEIS0012 NEIS0012	Ţ Ţ	porA porB NEIS1364 PorA_VR1 PorA_VR2 abcZ adk aroE aspA carB dhpS fumC	Ţ	

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. You can also use a usergroups.allow file. This file should contain the names of user groups, the members of which are allowed access. The file should contain one user group name 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 'User passwords' box in the admin section of the curator's index page. Click the 'Set passwords' link.



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

PubMLST			
Home > Organisms > Organ	sm > Neisseria typing > Change password		
Set user passv Passwords must be at least			2
Passwords			Action
User:	Jolley, Keith (keith)	~	SET PASSWORD
New password:			SETTASSWORD
Retype 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 => 'BORSI',
author => 'Keith Jolley',
                 affiliation => 'University of Oxford, UK',
                 email => 'keith.jolley@zoo.ox.ac.uk',
                 description => 'Perform BURST cluster analysis on query results query_
\leftrightarrow results',
                 category => 'Cluster',
                 buttontext => 'BURST',
                 menutext => 'BURST',
                module => 'BURST',
version => '1.0.0',
dbtype => 'isolates, sequences',
section => 'postquery',
order => 10,
                 system_flag => 'BURST',
                 input => 'query',
                 requires => 'mogrify',
                 min => 2,
                              => 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. If the system_flag value is set to 'no' then the plugin will be disabled even if the all_plugins attribute is set to 'yes'.

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 --database 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.

You are able to update the cache for a single scheme, or a list of schemes, and choose the method of update. For large schemes, such as cgMLST, a full refresh may take a long time, so you may wish to only perform this infrequently (perhaps once a week) with more regular 'daily' or 'daily_replace' updates. A full list of options available are shown by typing

```
update_scheme_caches.pl --help
NAME
    update_scheme_caches.pl - Update scheme field caches
SYNOPSIS
    update_scheme_caches.pl --database NAME [options]
OPTIONS
--database NAME
   Database configuration name.
--help
   This help page.
--method METHOD
    Update method - the following values are allowed:
    full: Completely recreate caches
    incremental: Only add values for records not in cache.
    daily: Only add values for records not in cache updated today.
    daily_replace: Refresh values only for records updated today.
--quiet
    Don't output progress messages.
--schemes SCHEMES
    Comma-separated list of scheme ids to use.
    If left empty, all schemes will be updated.
```

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.



You can also set cache_schemes="yes" in the system tag of config.xml to enable automatic refreshing of the caches (using the 'daily' method) 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 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.

See also:

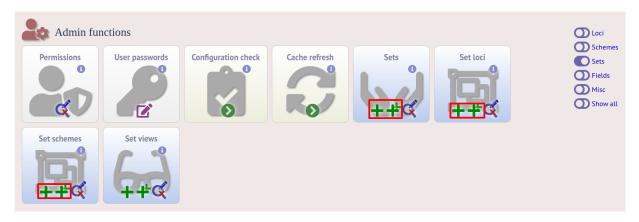
Sets (concept)

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. These functions are normally hidden, so you may need to click the 'Sets' toggle to display it.



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.

BIGSdb						
Home > Species ID > Ribosomal MLST genome						
Ribosomal MLST genome (Bacterial domain) database						
This database contains isolate and sequence data for publicly available genomes linked to allele definitions for the ribosomal proteins8:1005-15.	in genes. rMLST is	described in Jolley et al. 2012 Microbiology	/			
For academic non commercial use only. For full terms and conditions please see rMLST_licence.pdf. To discuss any other sort of use, innovation@isis.ox.ac.uk quoting reference 7895/MC.	including comme	rcial use licence, please contact				
Datasets						
This database contains multiple datasets.						
Please select: Bacterial domain Choose						
Query database		€ LOG OUT	+			
Search database		Z EXPORT	+			
Browse, search by any criteria, or enter list of attributes.		ANALYSIS				
		CUSTOMISE	+			
Search by combinations of loci This can include partial matches to find related isolates.			+			

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 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:

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.

5.12.4 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 Setting a site-wide users database

On large sites you may wish to employ a site-wide users database so that user details are kept in a single location and the user can log in to any database using the same credentials.

Once a *site-wide user database has been set up*, this can be defined within each client database as follows. From the curators' contents page, click the add (+) user databases link. This function is normally hidden, so you may need to click the 'Misc' toggle to display it.



Enter the user database details. You only need to enter the full database connection details if these are different from those set in db.conf. Press submit.

PubMLST			
Home > Organisms > Organism > Neis	eria isolates > Add new user database		
Add new user datal	ase		2
Please fill in the fields below - required	fields are marked with an exclamation mark (!).		and the second se
Record		Action	
id: 1		RESET SUB	
name: PubMLST	Site/domain name	KESET SUB	
dbase name:! pubmist_b			
curator:! Keith Jolley			
datestamp:! 2020-07-1			
list order:			
auto registration: Otrue O	-		
dbase host:	IP address of database host		
dbase port:	Network port accepting database connections		
dbase user:			
dbase password:			

Curators will need *specific permissions* set to be able to modify details in, or import users from a site-wide users database.

5.14 Adding new loci

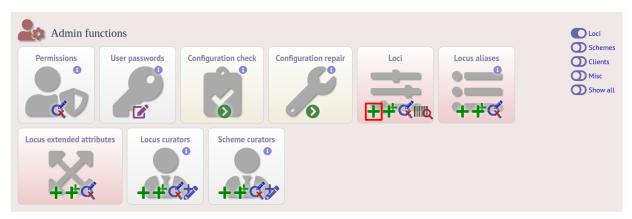
See also:

Loci (concept)

5.14.1 Sequence definition databases

Single locus

Click the add (+) loci link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Loci' toggle to display it.



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, number 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. This function is normally hidden, so you may need to click the 'Loci' toggle to display it.

Admin function	ons				Loci
Permissions	Jser passwords	Configuration check	Configuration repair	Locus aliases	Schemes Clients Misc Show all
Locus extended attributes		Scheme curato			

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

PubMLST		
Home > Organisms > Organism > Neisseria typing > Batch add locus records		
Batch insert loci		
Instructions 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. Templates Check the description of database fields for help with filling in the template. Upload		ł
Paste in tab-delimited text (Include a field header line).	Action	
	RESET	

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'.

5.14.2 Isolate databases

Single locus

Click the add (+) loci link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Loci' toggle to display it.

Admin functions				Loci	
Permissions	User passwords	Configuration check	Cache refresh	Locus aliases	 Schelles Fields Misc Show all
PCR reactions	PCR locus links	Nucleotide probes			

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

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Add new locus	
Add new locus	Show tools
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record	
idt idta type! DNA ~ allele id format! integer ~ length varies! true false coding sequence! isolate display! allele only ~ main display! true false query field! true false curator! keith Jolley (keith) date entered! 2020-07-19 formatted name: common name: locus type: locus type: intue false complete cds: intue false complete cds: intue false	Ē
pcr filter: true false Do NOT set to true unless you define PCR reactions linked to this locus. probe filter: true false Do NOT set to true unless you define probe sequences linked to this locus. introns: true false Do NOT set to true unless you define probe sequences linked to this locus. introns: true false De to NOT set to true unless you define probe sequences linked to this locus. introns: true false De to NOT set to true unless you define probe sequences linked to this locus. introns: true false De to NOT set to true unless you define probe sequences linked to this locus. introns: true false De to NOT set to true unless you define probe sequences linked to this locus. idbase name: pubmlst_bigsdb_neisseria_seqdef Name of the database holding allele sequences dbase password:	
submission template: Orue e false Include column in isolate submission template for this locus view: aliases: Action RESET SUBMIT	

- 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.
- 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.
- introns Set to true if locus may contain introns. This setting will only be available if BLAT is configured in bigsdb.conf.
 - 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_id Name of locus in seqdef database. This is usually the same as the id field.
 - 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)
- view Restrict this locus to only isolates contained in the specified database view. This option will only appear if the views attribute is set in the system tag. The view needs to have been defined in the database as a subset of the isolate table (usually filtered by the value of one or more of its fields).

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	
Home > Organisms > Organism > Neisseria isolates > Add new locus	
Add new locus	
	Show tools
Please fill in the fields below - required fields are marked with an exclamation mark (!).	ť
Record	
idt!	
data type: I DNA 🗸	
allele id format. integer >	
length varies:] Orue @ false	
coding sequence: etre O false	
isolate display: allele only ~	
main display:! Otrue 🖲 false	
query field:! ● true ○ false	
analysis:! end transformation of the second s	
curator:! Keith Jolley (keith)	
date entered: 2020-07-19	
datestamp:1 2020-07-19	
formatted name:	
common name: formatted common name:	
iornautee common name:	
allele id regex: length:	
complete cds: Otrue Ofalse	
genome position:	
match longest: Otrue O false	
reference sequence:	

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

Home > Organisms > Organism > Nelsseria Isolates > Add new locus Add new locus	Lide tools
Add new locus	Hide tools
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record id: id: id: id: id: id: id: i	non name, reference sequence, genome
reference sequence:	

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:

pcr filter:	⊖true ⊖false	Do NOT set to true unless you define PCR reaction	s linked to this	is locus.
probe filter:	⊖true ⊖false	Do NOT set to true unless you define probe seque	nces linked to 1	this locus.
introns:	⊖true ⊖false	Set to true if locus can contain introns.		
dbase name:	pubmlst_bigsdb_ne	isseria_seqdef		Name of the database holding allele sequences
dbase host:			IP address of	of database host
dbase port:		Network port accepting database connecti	ons	
dbase user:				
dbase password:				
dbase id:	PUT_LOCUS_NAM	IE_HERE 💀	Name of locu	us in seqdef database
description url:	/cgi-bin/bigsdb db=pubmlst_neis	o/bigsdb.pl? sseria_seqdef&page=locusInfo&locu:	S=PUT_LOCU	US_NAME_HERE
urt:	/cgi-bin/bigsdb db=pubmlst_neis _id=[?]	o/bigsdb.pl? sseria_seqdef&page=alleleInfo&locu	us=PUT_LOC	CUS_NAME_HERE&allele
submission template:	🔾 true 🙁 false	Include column in isolate submission template for	this locus	
view:		~		
aliases:				
Action				
RESET SUBMIT				

Complete the form and click 'Submit'.

Note: You can also pre-populate the dbase_name, dbase_url and url fields with boilerplate values by setting the default_seqdef_config and default_seqdef_dbase values in the system attribute of the config.xml file.

Batch adding

Click the batch add (++) loci link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Loci' toggle to display it.

Admin fur	nctions				Loci
Permissions	User passwords	Configuration check	Cache refresh	Locus aliases	 Chieffies Fields Misc Show all
PCR reactions	PCR locus links	Nucleotide probes			

Click the link to download an Excel template:

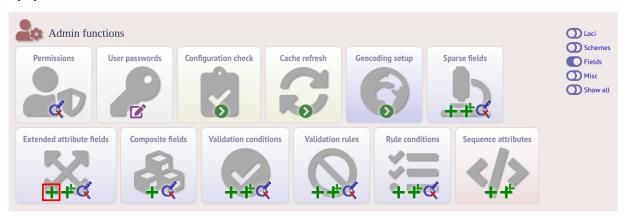
PubMLST		
Home > Organisms > Organism > Neisseria isolates > Batch add locus records		
Batch insert loci		8
Instructions This page allows you to upload locus data as tab-delimited text or copied from a spreadsheet. • Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish. • Enter aliases (alternative names) for your locus as a semi-colon (a) separated list. Templates The the description of database fields for help with filling in the template. Upload Paste in tab-delimited text (include a field header line).	Action RESET SUBMIT	

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'.

5.15 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. This function is normally hidden, so you may need to click the 'Fields' toggle to display it.



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

PubMLST			
Home > Organisms > 0	Organism > Neisseria isolates > Add new isolate field extended attribute		
Add new is	olate field extended attribute		2
Please fill in the fields	below - required fields are marked with an exclamation mark (!).	Action	- H
isolate field: attribute:		RESET	
datestamp:	Keith Jolley (keith) 2020-07-19		
value regex description			
urt			
length: field 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 (I) 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.

5.16 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.

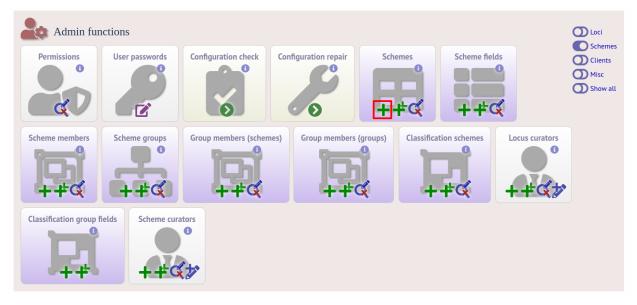
See also:

Schemes (concept)

5.16.1 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. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.

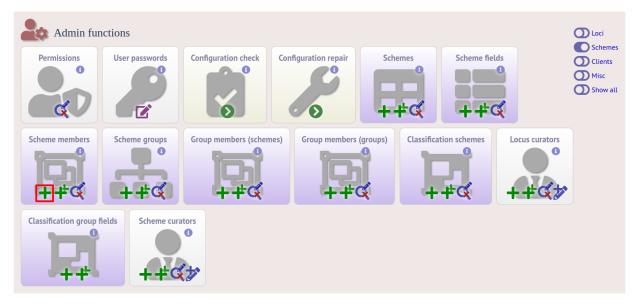


Fill in the scheme description in the web form. The next available scheme id number will be filled in already.

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.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new scheme	
Add new scheme	
Please fill in the fields below - required fields are marked with an exclamation mark (l). Record id: 1	Ŧ
name: MLST ① RESET SUBMIT	
curator:! Keith Jolley (keith) datestamp:! 2020-07-19 date entered: 2020-07-19	
description:	
display order:	
allow missing loci: Otrue Ofalse 👔 This is only relevant to schemes with primary key fields, e.g. MLST.	
max missing: ① Number of loci that are allowed to be missing for a profile to be defined.	
disable: Otrue O false Set to true to disable scheme. This can be overridden by user preference settings.	
no submissions: Otrue Ofalse Set to true to prevent submission of profiles of this scheme via the automated submission system.	
PubMed ids:	
links: (Format: URL/description)	

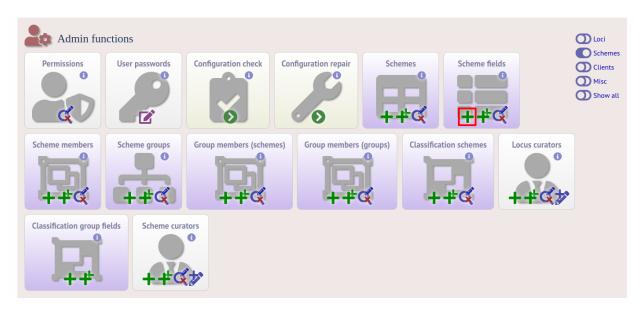
To add loci to the scheme, click the add (+) scheme members link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.



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.

PubMLST		
Home > Organisms > Organism > Neisseria typing > Add new scheme	member	
Add new scheme member		2
Please be aware that any modifications to the structure of this scheme v any 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 designation	ons, but
Please fill in the fields below - required fields are marked with an exclar	mation mark (!).	
Record	Action	
scheme id: MLST ~		
locus: abcZ	RESET	
curator:! Keith Jolley (keith)		
datestamp:! 2020-07-19		
field order: 1		

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



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

PubMLST		
Home > Organisms >	s > Organism > Neisseria typing > Add new scheme field	
Add new s	scheme field	20
	that any modifications to the structure of this scheme will result in the removal of all data from it. This is done to ensure data integrity. This does not affect allele designa	tions, but
any profiles will have	nave to be reloaded.	
Please fill in the field	fields below - required fields are marked with an exclamation mark (!).	
Record	Action	-
		ET.
scheme id:! field:!	DESET CURMIT	
type:		
	v! Otrue ● false ①	
	n:I Otrue e false 👔	
	I Keth Joley (keth)	
	n: 2020-07-19	
description:		
field order:	er:	
index:	ex: Otrue Ofalse 🚯	
value regex:	ex:	

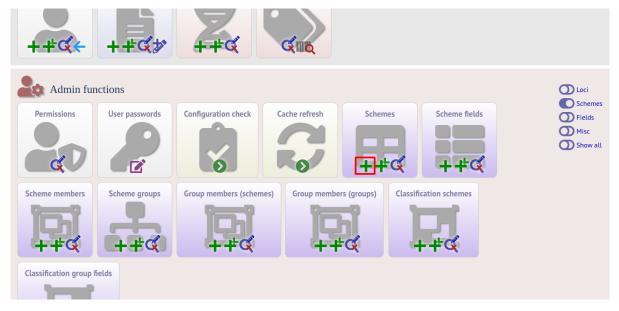
- scheme_id Dropdown box of scheme names.
 - Allowed: selection from list.
- field Field name.
 - Allowed: any value.
- type Format for values.
 - Allowed: text/integer/date.
- primary_key Set to true if field is the primary key. There can only be one primary key for a scheme.
 - Allowed: true/false.
- 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

5.16.2 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. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.



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

PubMLST		
Home > Organisms > Organism	m > Neisseria isolates > Add new scheme	
Add new scher	me	
Record id: name: isolate display: main display: query field: query status: analysis:	<pre>e true _ false () e true _ false () e true _ false () true _ false () e true _ false () 2020-07-19 2020-07-19</pre>	ŧ
dbase name: dbase host:		
dbase port:	•	
dbase user:		
dbase password:		
dbase id:		
view:	~ 0	
display order:	0	
	O true o false This is only relevant to schemes with primary key fields, e.g. MLST. experimental o in development please cite unpublished Use CTRL/SHIFT click to select or deselect values	
PubMed ids: links: (Format: URL description)		

- 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.
- recommended Sets whether the scheme will appear in a list of recommended schemes for use in some analysis plugins. Selecting this option makes the scheme easier to select when there are a lot of schemes defined. It should be used sparingly.
- quality_metric Sets whether the scheme can be used to help assess the quality of a genome assembly. For a well annotated genome it would be expected for all loci in the scheme to have an allele designated. The annotation status can be searched in an isolate query. This can be used in conjunction with the quality_metric_good and quality_metric_bad attributes that can be used to set the thresholds for what constitutes a good or bad annotation.
- 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_id Id of scheme in the sequence definition database.
 - Allowed: any integer.
- quality_metric_good threshold number of loci that must have allele designations for a genome annotation to be considered good for this scheme. If this isn't set then the number of loci in the scheme is used.
- quality_metric_bad threshold number of loci that must have allele designations below which a genome annotation is to be considered bad for this scheme. If this isn't set then the value used for quality_metric_good is used (or the number of scheme loci if this also is not set).
- view Restrict this scheme to only isolates contained in the specified database view. This option will only appear if the views attribute is set in the system tag. The view needs to have been defined in the database as a subset of the isolate table (usually filtered by the value of one or more of its fields).
- 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).

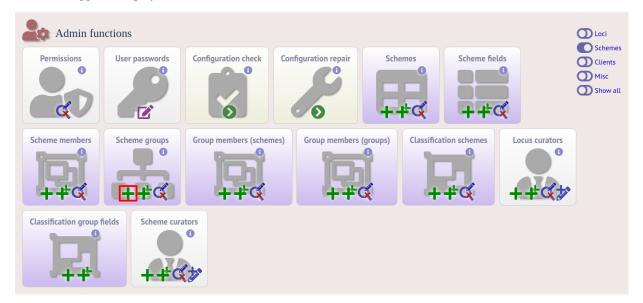
5.17 Organizing schemes into hierarchical groups

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.

BIGSdb		_				
Home > Orga	anisms > Or	ganism	> Neisseria typin	g > Downl	oad alleles	
Down	load a	llel	e sequen	ces		Help 🗹
Click withir		display	tical list All loci details of loci belc	·	hemes or groups o	of schemes - clicking a group folder will display the loci for all schemes within the group and any subgroups. Click the
	loci Strain typing Capsule Genetic Infor Genomic isla Lineage Schr Metabolism N. gonorrhoe Plasmids Typing Japhi Strateger Finetypin	mation inds emes eae AMR	-			
					Max length Full n	
abcZ adk		1077 841	Fixed: 433 bp Fixed: 465 bp	433 464	434 465	O. Harrison, K. Jolley 2020-06-23 O. Harrison, K. Jolley 2020-07-14
		1118		489	493	O. Harrison, K. Jolley 2020-01-14
			Fixed: 465 bp	463	467	O. Harrison, K. Jolley 2020-06-12
		1156		501	513	O. Harrison, K. Jolley 2020-07-03
			Fixed: 480 bp Fixed: 450 bp	480 447	503 489	O. Harrison, K. Jolley 2020-06-04 O. Harrison, K. Jolley 2020-06-12
pgm	X	11/2	. Acti 150 Up		107	

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. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.

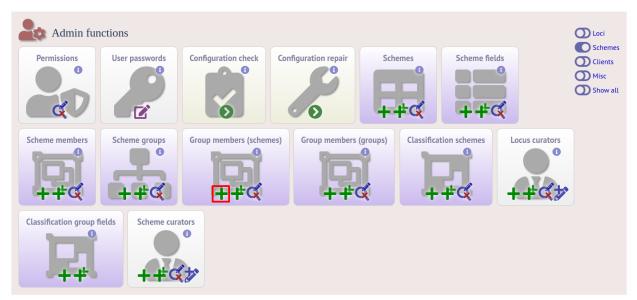


Enter a short name for the group - this will appear within drop-down list boxes and the hierarchical tree, so it needs to be fairly short.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new scheme group	
Add new scheme group	20
Please fill in the fields below - required fields are marked with an exclamation mark (!).	
Record Action	Ē.
id: 1 name: Typing	
curator: Ketth Jolley (keith)	
datestamp:/ 2020-07-19	
description:	
display order:	
seq query: Otrue Ofalse	

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 (+) group members (scheme) link.



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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new scheme group scheme member	
Add new scheme group scheme member	
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record group id:! Typing scheme id:! MLST curator:! Keith Jolley (keith) datestamp:! 2020-07-19	ľ

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

members link.

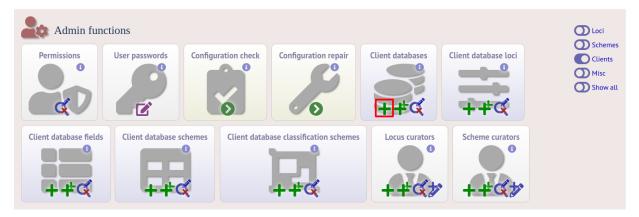
5.18 Setting up client databases

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. This function is normally hidden, so you may need to click the 'Clients' toggle to display it.



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

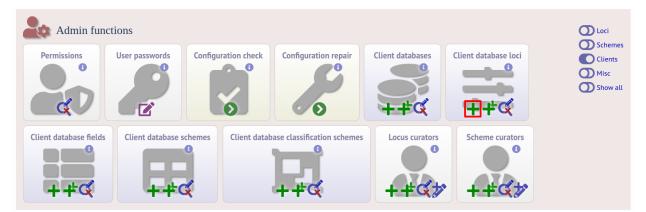
PubMLST			
Home > Organisms > Organ	ism > Neisseria typing > Add new client database		
Add new clie	nt database		20
Please fill in the fields belo	w - required fields are marked with an exclamation mark (!).		
Record			
id:! name:!	PubMLST isolates		et i
description:!	Contains data for a collection of isolates that diversity of <u>Neisseria</u> species. For every <u>allel</u> database there is at least one corresponding is isolate may be submitted to this database and c noted that it does not represent a population s	ic profile in the profiles Date deposited here. Any onsequently it should be	
dbase name:!	pubmlst_bigsdb_neisseria_isolates	Name of the database holding isolate data	
dbase config name:!	pubmlst_neisseria_isolates	Name of the database configuration	
	Keith Jolley (keith)		
datestamp:!	2020-07-19		
dbase host:		IP address of database host	
dbase port:	 Network port accepting database conr 	nections	
dbase user:		(i)	
dbase password:		(i)	
dbase view:		View of isolates table to use	
url:	/cgi-bin/bigsdb/bigsdb.pl	Web URL to database script	
Action			
RESET	МІТ		

- 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 seqdef 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.

5.18.1 Look up isolates with given allele

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



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	
Home > Organisms > Organism > Neisseria typing > Add new locus to client database definition	
Add new locus to client database definition	
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record client dbase id: 1) PubMLST isolates curator: abcZ curator: Keth Joley (keth) datestamp: 2020-07-19 locus alias: Action RESET SUBMIT	(f 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.

BIGSdb		
Home > Organisms > Organism	n > Neisseria typing > Allele information	
Allele informat	tion - abcZ: 5	Help 🔽
	abcZ	
length:	433	
	Sanger trace checked	
date entered:		
datestamp:	2009-11-11 Keith Jolley, University of Oxford, UK	
	Man-Suen Chan, University of Oxford	
Profiles contain		
Solate database	es	
	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 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 sample. 1578 isolates	

5.18.2 Look up isolates with a given scheme primary key

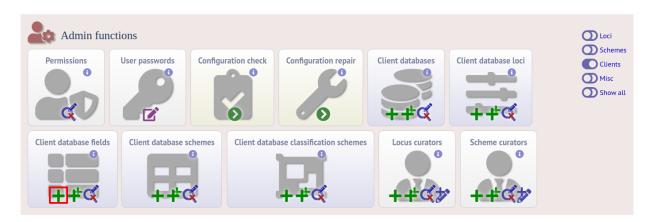
Setting this up is identical to setting up for alleles (see above) except you click on the add (+) client database schemes link and choose the scheme and client databases in the dropdown list boxes.

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.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Profile information	
Profile information for ST-11 (MLST)	🚹 Help 🔽 🚯
STabcZadkarocfumCgdhpdhCpgmclonal complex112343846ST-11 complex	
sender: Paula Kriz, Paula Kriz and Keith Jolley curator: Keith Jolley, University of Oxford, 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 profile one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it doe sample. 7313 isolates	

5.18.3 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.



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.

PubMLST		
Home > Organisms > Organism > Neisseria typing > Add new locus to client database isolate field	definition	
Add new locus to client database isolate field de	efinition	
Please fill in the fields below- required fields are marked with an exclamation mark (!).	Action RESET SUBMIT done.	ľ

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.

BIGSdb	
Home > Organisms > Organis	m > Neisseria typing > Allele information
Allele informa	tion - penA: 9
Provenance/m	eta data
locus: allele:	
	GACGGCGTTT TGCTGCCGGT CAGCTTTGAA AAACAGGCGG TTGCGCCGCA AGGCAAACGT ATATTTAAAG CATCGACCGC ACGTCAGGTG CGTGAGTTGA TGGTTTCTG AACCGAACCT GGCGGTACGG GTACGGCGGG GGCGGTAGAT GGCTTGGACG CGGGGCAAA AACCGGTACG GCGGTAAGT TGGTTAACGG TCGTTACGT GATTACAAAC ACGTTGCCAC TTTCATCGGT TTTGCCCGG CTGAAAATCC GCGGTGTAT GTGGCGGTAA CCATTGACGAG GCCGACTGCA AACGGTTACT ACGGCGGCGT AGTGACAGGT CCGGTCTTCA AACAAGTTAT GGGCGGTAGC TTGAGCGGTTTC TCCGACCAAA CCCTTGACCA AT
	Sanger trace checked
date entered:	2006-09-04
	2006-09-04
	Muhamed-Kheir Taha, National Reference Center for Meningococci, Institut Pasteur, France
curator: mutation F504L:	Keith Jolley, University of Oxford, UK
mutation A510V:	
mutation I515V:	yes
mutation H541N:	,
mutation I566V:	yes
Publication (1)	
Jolley KA, Kadlu	ez 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, bowski 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, (2007). Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis. <i>Antimicrob Agents Chemother</i> 51:2784-92
Isolate databas	ses
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. 979 isolates
🔗 Linked data	
penicillin_range:	>0.06 - 1 (intermediate) [n=143]; <=0.06 (susceptible) [n=1]; >1 (resistant) [n=1] PubMLST isolates

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_id: atpD
- 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_id: 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 Automated assignment of scheme profiles

It is not practical to define cgMLST profiles via the web interface. A script is provided in the scripts/automation directory of the BIGSdb package called define_profiles.pl that can be used to scan an isolate database and automatically define cgMLST profiles in the corresponding sequence definition database.

The script is run as follows:

```
define_profiles.pl --database <name> --scheme <scheme_id>
```

A full list of options can be found by typing:

```
define_profiles.pl --help
NAME
   define_profiles.pl - Define scheme profiles found in isolate database
SYNOPSIS
    define_profiles.pl --database NAME --scheme SCHEME_ID [options]
OPTIONS
--cache
   Update scheme field cache in isolate database.
--database NAME
   Database configuration name.
--help
   This help page.
--exclude_isolates LIST
   Comma-separated list of isolate ids to ignore.
--exclude_projects LIST
    Comma-separated list of projects whose isolates will be excluded.
--iqnore_multiple_hits
   Set allele designation to 'N' {f if} there are multiple designations set for
    a locus. The default \mathbf{is} to use the lowest allele value \mathbf{in} the profile
   definition.
--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).
--match_missing
    Treat missing loci as specific alleles rather than 'any'. This will
```

(continues on next page)

(continued from previous page)

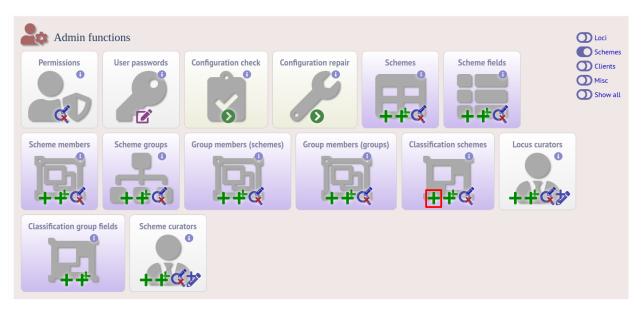
```
allow profiles for every isolate that has <= threshold of missing alleles
   to be defined but may result in some isolates having >1 ST.
--max ID
   Maximum isolate id.
--min ID
   Minimum isolate id.
--min_size SIZE
   Minimum size of seqbin (bp) - limit search to isolates with at least this
   much sequence.
--missing NUMBER
   Set the number of loci that are allowed to be missing in the profile. If
   the remote scheme does not allow missing loci then this number will be set
   to 0. Default=0.
--projects LIST
   Comma-separated list of project isolates to scan.
--scheme SCHEME ID
   Scheme id number.
--view VIEW
   Limit isolates searched to specified view.
```

5.21 Scheme profile clustering - setting up classification schemes

Classification groups are a way to cluster scheme profiles using a specified threshold of pairwise allelic mismatches. Any number of different classification schemes can sit on top of a standard scheme (such as cgMLST), allowing different similarity thresholds to be pre-determined. Currently, single-linkage clustering is supported whereby each member of a group must have no more than the specified number of allelic differences with at least one other member of the group.

5.21.1 Defining classification scheme in sequence definition database

Once a scheme has been defined, add a classification scheme by clicking the add classification schemes (+) link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.



Select the underlying scheme and enter a name for the classification scheme, the number of mismatches allowed in order to include a scheme profile in a group, and a description. An example name for such a scheme could be 'Nm_cgc_25' indicating that this is a classification scheme for *Neisseria meningitidis* core genome cluster with a threshold of 25 mismatches.

You can additionally choose whether a relative threshold is used to calculate the number of mismatches to account for missing loci in pairwise comparisons. In this case, in order to be grouped, the number of matching alleles must exceed:

(number of common loci x (total loci - defined threshold)) / total loci

rather than

total loci - defined threshold

when an absolute threshold is used.

As this threshold has to be calculated for each pairwise comparison, clustering using relative thresholds is slower than using an absolute value, and probably makes little real world difference.

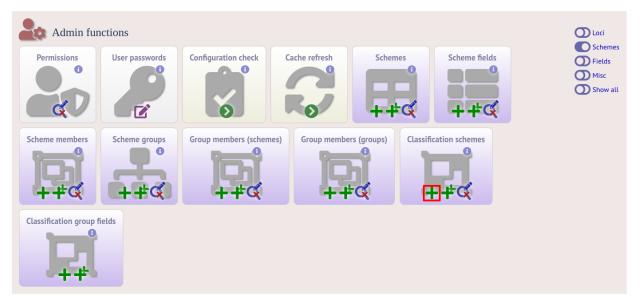
The status can be 'experimental' or 'stable'. The status of a scheme will be shown in the web interface to indicate that any groupings are subject to change and do not form part of the stable nomenclature.

Press 'Submit' to create the classification scheme.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new classification scheme	
Add new classification scheme	
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record id:! 2	†
scheme id: N. gonorrhoeae cgMLST v1.0	
name: Nm_cgc_25	
inclusion threshold: 25 Maximum number of different alleles alle	wed between profile and at least one group member profile.
use relative threshold: O true false Calculate threshold using ratio of shared/pre	sent in both profiles in pairwise comparison.
status: experimental 🗸	
curator:! Keith Jolley (keith)	
datestamp: 2020-07-19	
<pre>description: Single linkage clustering with each group membe allelic mismatches to at least one other member are ignored in comparisons.</pre>	
display order:	
Action	
RESET	

5.21.2 Defining classification scheme in isolate database

Duplicate the scheme definition from the sequence definition database. Click the add classification schemes (+) link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.



Enter the same details used in the sequence definition database. If a different id number is used in the isolate and sequence definition databases, you can set the seqdef id in the seqdef_cscheme_id field (the default is to use the same id).

You can also define a display order - this is an integer field on which the ordering of classification schemes is sorted when displayed in the isolate information page.

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Add new classification sch	me
Add new classification scheme	
Please fill in the fields below - required fields are marked with an exclamation man Record	к ().
id:! 2 scheme id:! N. meningitidis cgMLST v1.0 ❤	
name:! Nm_cgc_25	\odot
	f different alleles allowed between profile and at least one group member profile. g ratio of shared/present in both profiles in pairwise comparison.
use relative threshold:! O true • false () Calculate threshold usin status:! experimental	g ratio of shared/present in outripiontes in pariwise comparison.
curator:! Keith Jolley (keith)	
datestamp:! 2020-07-19	
description: Single linkage clustering with ea	ch group member having fewer than 25 e other member of the group. Missing loci
seqdef cscheme id: 2	ber defined in seqdef database
display order:	
Action	
RESET SUBMIT	

It is a good idea to *check the configuration*.

5.21.3 Clustering

Clustering is performed using the cluster.pl script found in the scripts/automation directory of the BIGSdb package. It should be run by the bigsdb user account (or any account with access to the databases).

Currently only single-linkage clustering is supported.

The script is run as follows from the command line:

cluster.pl --database <database configuration> --cscheme <classification scheme id>

A full list of options can be found by typing:

```
cluster.pl --help
NAME
    cluster.pl - Cluster cgMLST profiles using classification groups.
SYNOPSIS
    cluster.pl --database NAME --cscheme_id SCHEME_ID [options]
OPTIONS
--cscheme CLASSIFICATION_SCHEME_ID
    Classification scheme id number.
--database NAME
    Database configuration name.
--help
    This help page.
```

(continues on next page)

(continued from previous page)

```
--reset
Remove all groups and profiles currently defined for classification group.
```

Note: Note that for classification schemes to be accessible within the isolate database, *scheme cache tables* must be generated and kept up-to-date.

Where an isolate has been clustered in to a group with other isolates, this information is available in the *isolate information page*.

 Publications (2) Feavers IM, Gray SJ, Urwin R, Russell JE, Bygraves JA, Kaczmarski EB, Maiden MC (1999). Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak. <i>J Clin Microbiol</i> 37:5883-7 12 isolates Jolley KA, Hill DM, Bratcher HB, Harrison OB, Feavers IM, Parkhill J, Maiden MC (2012). Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid Web-based analysis methods. <i>J Clin Microbiol</i> 50:3046-53 12 isolates Sequence bin contigs: 259 N50 contig number: 18 N95 contig number: 79 N50 length (L50): 38,364 N95 length (L95): 4,593
Web-based analysis methods. J Clin Microbiol 50:3046-53 12 isolates Sequence bin contigs: 259 N50 contig number: 18 N95 contig number: 79
contigs: 259 N50 contig number: 18 N95 contig number: 79
wide wide <thwide< th=""> wide wide <thw< th=""></thw<></thwide<>
Show sequence bin
Experimental schemes are subject to change and are not a stable part of the nomenclature.
Classification scheme Underlying scheme Clustering method Mismatch threshold Status Group Nm cac 200 N. meningitidis coMLST v1.0 Single-linkage 200 experimental group; 17 (1007 isolates)
Nm_cgc_200 (i) N. meningitidis cgMLST v1.0 Single-linkage 200 experimental group: 17 (1007 isolates) Nm_cgc_100 (i) N. meningitidis cgMLST v1.0 Single-linkage 100 experimental group: 38 (735 isolates)
Nm cqc 50 (1) N. meningtidis cgMLST v1.0 Single-linkage 50 experimental group 50 (15) Solates)
Nm_cgc_25 ① N. meningitidis cgMLST v1.0 Single-linkage 25 experimental group: 45 (4 isolates)
Schemes and loci

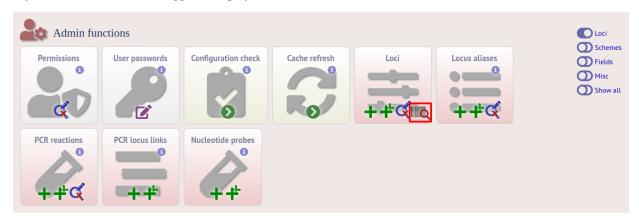
Clicking the hyperlinks will take you to a table containing matching isolates, from where standard analyses can be performed.

BIGSdb	
Home > Organisms > Organism > Neisseria isolates > Search or browse database	
Search or browse database	🔝 Help 🕜 🟮 🗡
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.	
Isolate provenance fields Allele designations/scheme fields	
id v = v Enter value I + (j Nm_cgc_25 group = v 45	+ ()
Display/sort options Action	
Order by: id vascending v Display: 25 v records per page ()	
Your projects Bookmark query Select project Add these records 2020-07-19:1 Add bookmark	
Isolate fields 🕧 MLST Finetyping antigens	
id isolate aliases country year disease species capsule group ST clonal complex PorA_VR1 PorA_VR2 FetA_VR 662 2837 M97/25208 UK 1997 invasive (unspecified/other) Neisseria meningitidis C 50 ST-11 complex 5-1 10-4 F3-6	
670 2840 M97/252535 UK 1997 invasive (unspecified/other) Neisseria meningitidis C 50 ST-11 complex 5-1 10-4 F3-6	
671 2844 M97/252781 UK 1997 invasive (unspecified/other) Neisseria meningitidis C 50 ST-11 complex 5-1 10-4 F3-6 672 2847 M97/252943 UK 1997 invasive (unspecified/other) Neisseria meningitidis C 50 ST-11 complex 5-1 10-4 F3-6	
Analysis tools	
Breakdown: Fields Two Field Combinations Polymorphic sites Publications Sequence bin	
Analysis: BURST Codons Gene Presence Genome Comparator BLAST rMLST species id PCR	
Export: Dataset Contigs Sequences	
Third party: GrapeTree ITOL PhyloViz Microreact	

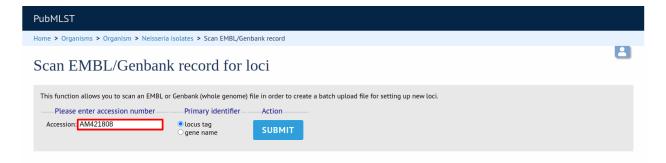
5.22 Defining new loci based on annotated reference genome

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.

Click 'Database scan' within the 'Loci' group on the curator's contents page. This function is normally hidden, so you may need to click the 'Loci' toggle to display it.



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



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

PubMLST		
Home > Organisms > Or	ganism > Neisseria isolates > Scan EMBL/Genbank record	
, ,		8
Scan EMBL	/Genbank record for loci	_
This function allows yo	to scan an EMBL or Genbank (whole genome) file in order to create a batch upload file for setting up new loci.	
Please enter acce	ssion number Primary identifier Action	
Accession: AM42180	O gene name SUBMIT	
Download tabl	e Download alleles	
	uitable for batch	
i x`	pload of loci.	
Annotation info	rmation	
acces	ion: AM421808	
	ion: 1	
	ype: dna gth: 2194961	
descrip	ion: Neisseria meningitidis serogroup C FAM18 complete genome.	
	ons: 1975	
Coding sequence	es	
Locus Aliases	Product	Length 924
NMC0001 lpxC; envA	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase	
NMC0002 pilS1	pilin (ragment)	291
NMC0003 pilS2 NMC0004 fbp	truncated pilin peptidyl-prolyl cis-trans isomerase	366 330
NMC0005	putative membrane protein	219
NMC0006	putative glycerate dehydrogenase	954
NMC0007 metG	methionyl-tRNA synthetase	2058 1839
NMC0008 glmS NMC0009	glucosaminefructose-6-phosphate aminotransferase [isomerizing] putative lipoprotein	519
NMC0010 gna33	outer membrane lipoprotein Gna33 ①	1326
NMC0011	putative integral membrane protein	840
NMC0012	putative lipoprotein	1167
NMC0013	possible membrane protein	1266

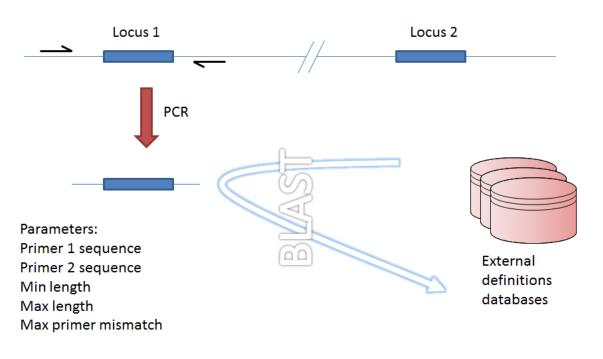
Tab-delimited text and Excel format files will be created to be used as the basis for upload files for the sequence definition and isolate databases. Batch sequence files, in text and Excel formats, are also created for defining the first allele once the locus has been set up in the sequence definition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

5.23 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.

5.23.1 Filtering by in silico PCR

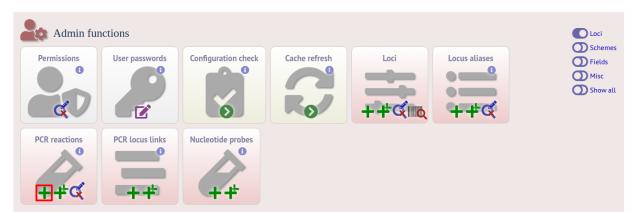
Provided a locus can be predicted to be specifically amplifed by a PCR reaction, the genome can be filtered to only look at regions predicted to fall within amplification products of one or more PCR reactions. Since this is *in silico* we don't need to worry about problems such as sequence secondary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



Locus 1 and locus 2 share allele pool

Fig. 1: Genome filtering by *in silico* PCR.

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. This function is normally hidden, so you may need to click the 'Loci' toggle to display it.

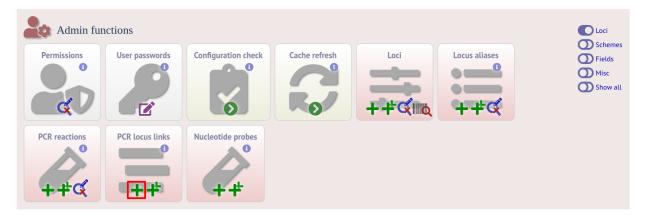


In the resulting web form you can enter values for your two primer sequences (which can be any length), the minimum and maximum lengths of reaction products you wish to consider and a value for the allowed number of mismatches per primer.

PubMLST						
Home > Organisms > Organism > Neisseria isolates > Add new PCR reaction						
Add new PCR r	reaction					
Please fill in the fields below - r	equired fields are marked with an exclamation mark (!).					
Record	A	Action				
id:!	2	RESET SUBM	-			
description:		SUBM				
primer1:!						
primer2:!						
curator:	Keith Jolley (keith)					
datestamp:!	2020-07-19					
min length:						
max length:						
max primer mismatch:	 Maximum sequence mismatch per primer 					

- 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.



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.

5.23.2 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.

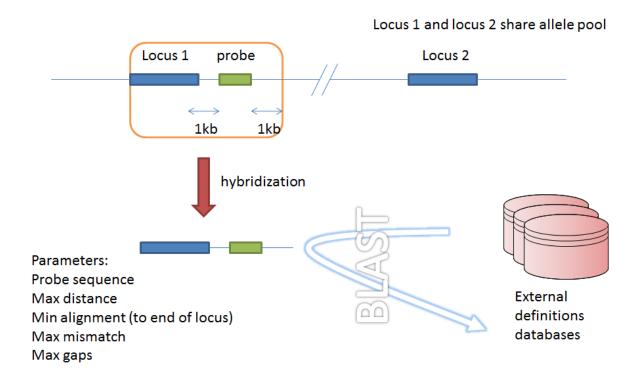
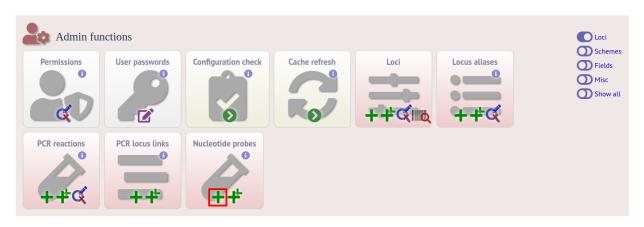
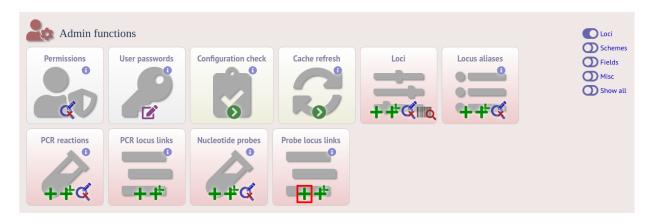


Fig. 2: Filtering by in silico hybridization

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.
 - 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.

5.24 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.



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

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Query or update isol	es
Query or update isolates	Help 🗹 🕄 🧨
Enter search criteria or leave blank to browse all records. Modify form parar	ters to filter or enter a list of values.
Isolate provenance fields	Display/sort options
isolate v = v MC58	+ (i) Order by: id vascending v Display: 25 v records per page (i)
Action SEARCH	
1 record returned.	
DeleteTag scanningProjects	Your projects
Delete ALL Scan Select project	Link Select project Add these records
Bookmark query	
2020-07-19:1 Add bookmark	
Delete Update Sequence bin New version disolate allases	Isolate fields () MLST Finetyping antigens country year disease species capsule group ST Clonal complex PorA_VR1 PorA_VR2 FetA_VR UK 1983 Neisseria meningitidis B 74 ST-32 complex 7 // 16-2 // F1-5 //

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

1	Publications (7	7)
		itey SD, Vernikos GS, Quail MA, Cherevach I, White B, Parkhill J, Maiden MC (2010). Independent evolution of the core and accessory gene sets in the genus Neisseria: from the genome of Neisseria lactamica isolate 020-06. <i>BMC Genomics</i> 11:552 78 isolates
		y KA, Earle SG, Corton C, Bentley SD, Parkhill J, Maiden MC (2012). A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species s Neisseria. <i>Microbiology</i> 158 :1570-80 55 isolates
	Bennett JS, Jolle	y KA, Maiden MC (2013). Genome sequence analyses show that Neisseria oralis is the same species as 'Neisseria mucosa var. heidelbergensis'. Int J Syst Evol Microbiol
		\sim
Z	Sequence bin	
	contigs:	
	length: loci tagged:	2272360 bp 2,230
	Show sequence bin	

Click the 'Renumber' button:

equence	Sequencing method	Original designation	Length	Comments	Locus	Start	End	Direction	Annotation	Renumber (
1	Sanger		2272360	whole genome	NEIS2139	7	498	←	EMBL GBK Renumber				
				NEIS2140	502	897	←	CIMBE ODK	Kentamber				
				NEIS2141	918	2312	←						
					NEIS2142	2517	3161	←					
					NEIS2143	3158	3511	←					
					NEIS2144	3635	4117	\rightarrow					
					NEIS2145	4311	4961	\rightarrow					
					NEIS2146	4958	5875	\rightarrow					
							NEIS2147	5936	6214	\rightarrow			
						NEIS2148 (pgk)	6281	7492	←				
					NEIS2149	7573	8826	←					
						tRNA-lys	9197	9272	←				
					NEIS2150	9346	10317	←					
					NEIS2151	10350	10811	←					
					NEIS2152 (kdtA)	10840	12177	←					
					NEIS2153	12174	13622	←					
					NEIS2979	13848	13964	←					
						NEIS0001 (lpxC)	15221	16144	←				
					NEIS0210 (pilE)	17229	17741	←					
					NEISp0210	17232	17741	←					
					pilS	18127	23903	←					
					NEIS0004 (fbp)	23904	24233	←					
					NEIS0005	24311	24529	←					

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

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Renumber locus genome positions	

Renumber locus genome positions based on tagged sequences

ou have selected	to renumber the genome positions se	t in the locus table based
Option		Action
Remove positi	RENUMBER	
ne following desi	gnations will be made:	
Locus	Existing genome position +	New genome position +
NEIS2139		7
NEIS2140		502
NEIS2141		918
NEIS2142		2517
NEIS2143		3158
NEIS2144		3635
NEIS2145		4311
NEIS2146		4958
NEIS2147		5936
NEIS2148		6281
NEIS2149		7573
tRNA-lys		9197
NEIS2150		9346
NEIS2151		10350
NEIS2152		10840

5.25 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.

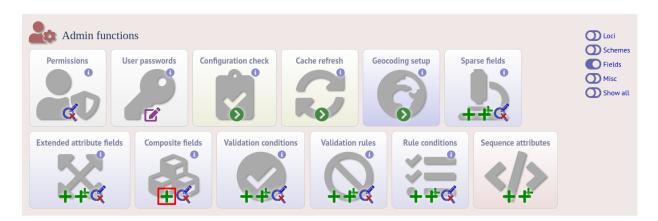
One example of a composite field is used in the Neisseria PubMLST database which has a strain designation composite field made up of serogroup, PorA VR1 and VR2, FetA VR, ST and clonal complex designations in the format:

[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'.

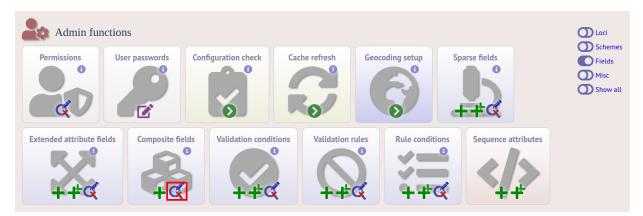
Composite fields can be added to the database by clicking the add (+) composite fields link on the curator's main page. This function is normally hidden, so you may need to click the 'Fields' toggle to display it.



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					
Home > Organisms > Organism > Neisseria isolates > Add new composite field					
Add new composite field					
Please fill in the fields below - required fields are marked with an exclamation mark (!)	L				
Record		Action			
id:! strain_designation	name of the field as it will appear in the web interface	RESET SUBMIT			
position after: isolate field present in the isolate table		RESET			
main display:! O true 🖲 false Sets whether to display field in isolate query results table (can be overridden by user preference).					
curator:! Keith Jolley (keith)					
datestamp:! 2020-07-19					

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.



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

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Update or delete composite field	
Update or delete composite field	2
1 composite field defined.	
Detete Update field name position after main display definition missing data X Image: Strain_designation isolate false [capsule_group]: P1.[PorA_VR1],[PorA_VR2]: [FetA_VR]: ST-[scheme 1:ST] ([scheme 1:clonal_complex)] ND: P1.ND,ND: F-ND: ST-[scheme 1:ST]	T-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.



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'.

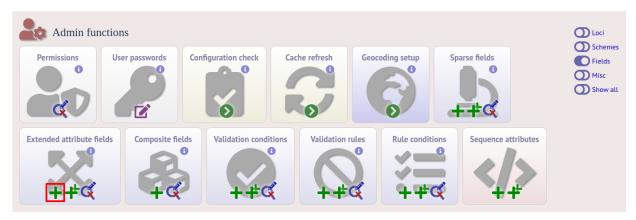
You can also define text to be used for when the field value is missing, e.g. 'ND'.

5.26 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. This function is normally hidden, so you may need to click the 'Fields' toggle to display it.

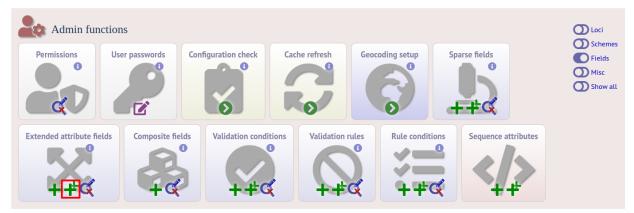


Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

- 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). This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



Download the Excel template:

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Batch add isolate field extended attribute records	
Batch insert isolate field extended attributes	
Instructions This page allows you to upload isolate field extended attribute data as tab-delimited text or copied from a spreadshe Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish. Templates Check the description of database fields for help with filling in the template. Upload Paste in tab-delimited text (Include a field header line).	

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

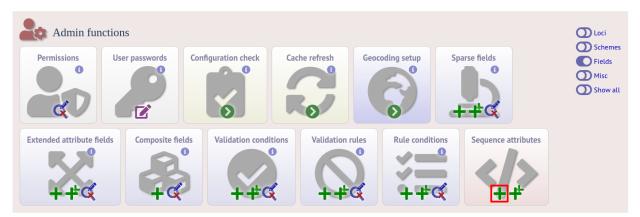
isolate_field	attribute	field_value	value
country	continent	Afghanistan	Asia
country	continent	Albania	Europe
country	continent	Algeria	Africa
country	continent	Andorra	Europe
country	continent	Angola	Africa

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

5.27 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. This function is normally hidden, so you may need to click the 'Fields' toggle to display it.



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'.

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Add new sequence attribute	
Add new sequence attribute	2 0
Please fill in the fields below - required fields are marked with an exclamation mark (!).	
- Record Action	
key:! read_length type! integer > RESET SUBMIT	
curator:! Keth Jolley (Keith)	
datestamp:! 2020-07-19 description:	

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

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Add new sequences	
Upload sequences	
This page allows you to upload sequence data for a specific isolate record in FASTA format. 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 (l). Paste in sequences in FASTA format:	X
Attributes Options isolate id: ! Enter isolate id sender: ! Select sender method: Image: Sender: ! Don't insert sequences containing only homopolymers. Link to experiment: Alternatively upload FASTA file or enter Genbank accession Select FASTA file: Choose file No file chosen Action RESET SUBMIT 	

5.28 Checking external database configuration settings

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



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. By default, only loci which have an issue will be displayed but you can click the 'show all loci' link to display them all.

Configura		Configuration	on check - Neis	seria isolate	s						
Configura		-									
Jungung	ation check - Neis	sseria	isolate	S							
	nion check Tiels	osciia	isolute	0							
DestaurSOI											
PostgreSQL											
BIGSdb requires P	g 9.5 or higher. You are running versio	on 10.12. 🗸									
Helper appli	cations										
Program	Path	Installed Ex	ecutable								
EMBOSS infoaligr		~	~								
EMBOSS sixpack		×	×								
EMBOSS stretche		~	×								
GrapeTree	/usr/local/GrapeTree/grapetree.py	× .									
blastn	/usr/bin/blastn	~	×								
blastp	/usr/bin/blastp	~	*								
blastx blat	/usr/bin/blastx /usr/local/bin/blat	ž	¥								
clustalw	/usr/bin/clustalw	~	¥								
ipcress	/usr/bin/ipcress	~	~								
mafft	/usr/bin/mafft	× .	×								
makeblastdb	/usr/bin/makeblastdb	 Image: A second s	~								
mogrify	/usr/bin/mogrify	 Image: A second s	×								
muscle	/usr/bin/muscle	~	~								
tblastx	/usr/bin/tblastx	~	~								
Locus datab	ases (only showing loci	with pot	ontial pro	hloms -	show all	loci)					
Locus datab	ases (only showing loci	with pot	ential pro	blems -	show all	loci)					
Locus	Database	Host	Port Id field	value Datai	ase accessibl	e Sequence qu	ery Sequen		1		
Locus NEIS0895 (parA)	Database pubmlst_bigsdb_neisseria_seqdef_zc	Host oo-aberlour	Port Id field	value Datai 895	ase accessibl	e Sequence qu	ery Sequen	×	1		
Locus NEIS0895 (parA) NEIS0903 (opaD)	Database pubmlst_bigsdb_neisseria_seqdef_zc pubmlst_bigsdb_neisseria_seqdef_zc	Host oo-aberlour oo-aberlour	Port Id field 5432 NEISO 5432 NEISO	value Datai 895 903	ase accessibl	e Sequence qu	ery Sequen	× ×	ı		
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454	Database pubmlst_bigsdb_neisseria_seqdef zc pubmlst_bigsdb_neisseria_seqdef zc pubmlst_bigsdb_neisseria_seqdef zc	Host oo-aberlour oo-aberlour oo-aberlour	Port Id field 5432 NEISO 5432 NEISO 5432 NEISO 5432 NEISO	value Data 895 903 454	oase accessibl	e Sequence qu	ery Sequen	x x x	1		
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC)	Database pubmlst_bigsdb_neisseria_seqdef_zc pubmlst_bigsdb_neisseria_seqdef_zc pubmlst_bigsdb_neisseria_seqdef_zc	Host oo-aberlour oo-aberlour oo-aberlour oo-aberlour	Port Id field 5432 NEISO 5432 NEISO 5432 NEISO 5432 NEISO 5432 NEISI 5432 NEISI	value Data 895 903 454 551	oase accessibl	e Sequence qu	ery Sequen	× ×	1		
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454	Database pubmlst_bigsdb_neisseria_seqdef zc pubmlst_bigsdb_neisseria_seqdef zc pubmlst_bigsdb_neisseria_seqdef zc	Host oo-aberlour oo-aberlour oo-aberlour oo-aberlour oo-aberlour	Port Id field 5432 NEISO 5432 NEISO 5432 NEISO 5432 NEISI 5432 NEISI 5432 NEISI 5432 NEISI 5432 NEISI 5432 NEISI	value Datal 895 903 454 551 013	oase accessibl	e Sequence qu	ery Sequen	× × × × ×	1		

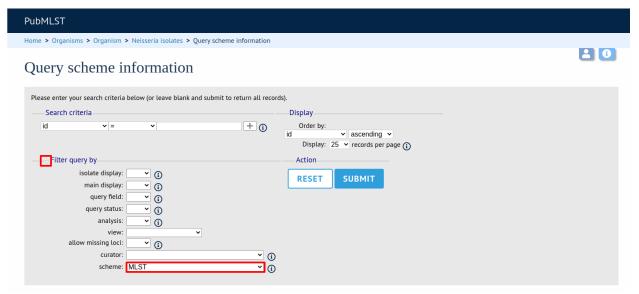
Any problems will be highlighted with a red X.

5.29 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 query link (Update or delete) next to schemes in the curator's interface. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.



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



Click the button 'Export configuration/data'.

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Query scheme information	
Query scheme information	
Please enter your search criteria below (or leave blank and submit to return all records). Search criteria id Order by id Display Display	: v ascending v : 25 v records per page ①
Filter query by Action isolate display:	SUBMIT
1 record returned. — Delete — Database configuration — Delete ALL Export configuration/data	
Delete Update id name dbase name host port id isolate display* c X 1 MLST pubmlst_bigsdb_neisseria_seqdef 1 Image: Comparison of the second seco	main isplay* query field* query status* query analysis* display view allow order curator datestamp datestamp date V V V 1 Keith Jolley 2019-07- 26 2009- 11-12

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_
→order allow_missing_loci curator datestamp date_entered
1 MLST pubmlst_bigsdb_neisseria_seqdef
                                          mv_scheme_1 1 1 1 1 1 1 1
→ 2 2012-03-22 2009-11-12
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
1 fumC
        4 2 2009-11-12
        5 2 2009-11-12
1 gdh
        6 2 2009-11-12
1 pdhC
        7 2 2009-11-12
1 pgm
scheme_fields
_____
scheme_id field type primary_key description field_order url isolate_display

->main_display query_field dropdown curator datestamp

1 ST integer 1 1 /cgi-bin/bigsdb/bigsdb.pl?page=profileInfo&db=pubmlst_
→neisseria_seqdef&scheme_id=1&profile_id=[?] 1 1 1 0 2 2010-01-20
1 clonal_complex text 0 2 1 1 1 1 2 2009-11-16
```

5.30 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/maintenance 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).
```

5.31 BLAST caches

Sequence definition databases cache any BLAST databases that they create in order to perform sequence queries. These caches can be found in subdirectories named with the database name in the temp directory defined by the secure_tmp_dir attribute in bigsdb.conf, e.g. /var/tmp/pubmlst_bigsdb_neisseria_seqdef.

These BLAST databases will be marked stale if new alleles are added to the BIGSdb database for any locus covered by the cache. A cache marked stale will be recreated the next time a matching sequence query needs to use it. BLAST databases will also be marked stale if they are older than the cache_days setting in bigsdb.conf (default = 7 days).

It is possible to also manually create and refresh these caches using the update_blast_caches.pl script found in the scripts/maintenance directory.

A full list of options can be found by typing:

```
update_blast_caches.pl --help
NAME
    update_cached_blast_dbs.pl - Refresh BLAST database caches
SYNOPSIS
   update_cached_blast_dbs.pl --database DB_CONFIG [options]
OPTIONS
--all_loci
   Refresh or create cache for all loci.
--database DATABASE CONFIG
   Database configuration name.
--delete_all
   Remove all cache files.
--delete_old
   Remove cache files older than the cache_days setting in bigsdb.conf or
   that have been marked stale.
--delete_single_locus
   Remove caches containing only one locus. There can be many of these and
   they can clutter the cache directory. They are generally quick to recreate
   when needed.
--help
    This help page.
--quiet
   Only show errors.
--refresh
   Refresh existing caches.
--scheme SCHEME_ID
    Refresh or create cache for specified scheme.
```

5.32 Config-specific file downloads

You can make files available on a static website but restrict their access only to users who can authenticate for access to the current database configuration.

This can be done by adding a file called download_files.conf to the database configuration directory within /etc/bigsdb/dbases. This file consists of three columns in tab-delimited format:

- The full path of the file in the file system
- Label which will be used to hyperlink to the file
- A description of the file
- The file type (docx, html, gif,jpg, pdf, png, tar, tar, xlsx currently supported)

The files can be downloaded directly from a BIGSdb URL: /cgi-bin/bigsdb.pl?db=CONFIG&page=downloadFiles&file=LABEL (where CONFIG is the database config name and LABEL is the label used in the download_files.conf file. These URLs can be used as standard links within a web page.

You can also list all available files with the URL: /cgi-bin/bigsdb.pl?db=CONFIG&page=downloadFiles

Navigating to these links will prompt the user to log-in if they are not already (if the database config requires this).

CHAPTER 6

Curator's guide

Please note that links displayed within the curation interface will vary depending on database contents and the permissions of the curator. Some infrequently used links are usually hidden by default. These can be enabled by clicking the 'Show all' toggle switch. The admin section has feature- specific toggles as well as a 'Show all' toggle.



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.



Enter the user's details in to the form.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Add new user	
nome - organisms - neissena spy neissena isotates - Aug nen aser	
	Help 🕜 🚯
Add new user	
Please fill in the fields below - required fields are marked with an exclamation mark (!).	
Record	Action T
id:1 548	
user name: jbloggs	RESET SUBMIT
surname:! Bloggs	
first name: I Joe	
email:1 joe.bloggs@zoo.ox.ac.uk	
affiliation: University of Oxford, UK	
status:! user	
date entered:! 2020-07-22	
datestamp: 2020-07-22	
curator:! Keith Jolley (Keith) submission emails: Otrue false Receive new submission E-mails (curators only)	
\sim	
account request emails: Otrue • false () Receive new account request E-mails (curators only)	
private quota: 0 User must be either a submitter, curator, or admin to upload private records	

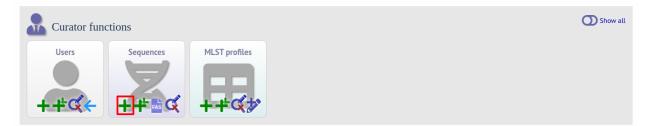
Normally the status should be set as 'user'. Only admins and curators with special permissions can create users with a status of curator or admin.

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



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.

You can also choose whether to designate the sequence as a type allele or not. Type alleles can be used to constrain the sequence search space when defining new alleles using the *web-based scanner* or *offline auto allele definer*.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new allele sequence	
Add new allele sequence	8
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record Locust: abcZ allele id: 1079 sequence: TTTGATACCGTTGGCGAAGGTTTGGGTAAAATTCGCGATTTATTGCGCCAGGTCAGCGAGGAGGCTGGGAA AAACCGTTCGGCGAGGTCAGCGGAAGCTCAGCGGTAGCGGAAAAATCGGCAGCTGGGAA GCTGGGATCGGCAGTCAGCGCGATCGGTCGGCCAGGCTTGGGTGCGCAGAGCCGCGCGGGTCAGGAGCGCGCGGGATAG GCTGGGAAAAAGCGTTCGGCGAAGCTTGGGTGCGACAGACCTGGCGGAGACCGGACGCAGCGATGGGAACCTGGCGGAGCGGCTTGGAGCGGCGCGGGATAGCGGAACCGGCGGGTTGGGGAAACCTGCCGCAGGCTTGGGGGAACCTGGCGGGGCGGGAGACCGGCGGGGATGGCGGAAGACCGGCGGGTTGGGGGAAACCTGGCGGGGTTGGGTGCGGCAGGCGCTGGGATGGCGGCAGCGGGGGGGG	* 1
status:! Sanger trace checked sender:! Jolley, Keith (keith) curator:! Keith Jolley (keith) date entered:! 2020-07-22 datestemp:! 2020-07-22 type allele: Otrue of alse New allele searches can be constrained to use just type alleles in comparisons comments: Comments: Comments:	
Flags: alternative start codon atypical contains IS element downstream fusion frameshift	
PubMed ids:	
ENA ids:	
Genbank ids: Override sequence similarity check Override sequence length check	

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

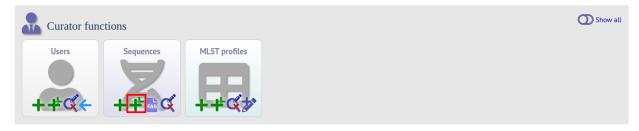
Sequences can also be associated with PubMed, ENA or Genbank id numbers by entering these as lists (one value per line) in the appropriate form box.

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 link on the curator's main page.



Download a template Excel file from the following page.

PubMLST			
Home > Organisms > Organism > Neisseria typing > Batch add new allele sequence records			
Batch insert sequences			Help 🗹
 Instructions 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 in WGS: automatically checked, 'unchecked'. Sequence flags can be added as a semi-colon () separated list. Sequence flags can be added as a semi-colon () separated list. Reload page specific for locus: Select Reload 	(BIGSdb)', 'WGS:		t (BIGSdb); 'WGS: visually checked',
Upload			
Please select the sender from the list below:			
Select sender Value will be overridden if you include a sender field in your pasted of generating or duplicate sequences		and no internal st	top codons. Existing sequences are
Paste in tab-delimited text (include a field header line).	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.

The status can be either: 'Sanger trace checked', 'WGS: manual extract (BIGSdb)', 'WGS: automated extract (BIGSdb)', 'WGS: visually checked', 'WGS: automatically checked' or 'unchecked'. See full explanations for these in the *single allele upload* section.

The 'type_allele' field is boolean (true/false) and specifies if the sequence should be considered as a type allele. These can be used to constrain the sequence search space when defining new alleles using the *web-based scanner* or *offline auto allele definer*.

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.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Batch add new allele sequence records	
Batch insert sequences	Hetp 🗹
Instructions In	act (BIGSdb)', WGS: automated extract (BIGSdb)', WGS: visually checked', nsert page specific to that locus: ted data. p codon at the ends and no internal stop codons. Existing sequences are Action RESET SUBMIT

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'.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Batch add new allele sequence records	
Batch insert sequences	Letp 🗹
Import status Sender: Keith Jolley No obvious problems identified so far.	
Data to be imported	
The following table shows your data. Any field with red text has a problem and needs to be checked. <i>Note: valid sequence flags are displayed with a red background not red text</i> locus allele_ld sequence sequence type allele sender curator [date_entered] detestamp [comm	
abcz 1079 TTTGATACCGTTGCCGAAGG GCGGATTGTTGAACTTGACC WGS: automated extract (BIGSdb) 2 2 2020-07-22 200-07-22 2020-07-22 2020-07-22 2020-07-	

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Batch add new allele sequence records	
Batch insert sequences	Hetp 🗹
Primary key Problem(s) locus: abcZ; allele_id: 1079 Sequence contains non nucleotide (AICIGIT) characters.	
Data to be imported The following table shows your data. Any field with red text has a problem and needs to be checked. Note: valid sequence flag: locus allele Id sequence status type allele	s are displayed with a red background not red text.
abcZ 1079 TTTGATACCGTTGCCGAAGG GCGGATTGTTGAACTTGACC WGS: automated extract (BIGSdb) abcZ 1080 TTTGATACCCTTGCCGAAGG GCGGATTGTTGAACTTGACC WGS: automated extract (BIGSdb)	2 2 2020-07-22 2020-07-22 2 2 2020-07-22 2020-07-22

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.

Curator fund	ctions		Show all
Users	Sequences + + + K	MLST profiles	

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

PubMLST	
ome > Organisms > Organism > Neisseria typing > Batch insert sequences	
Batch insert sequences	
This page allows 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 (loci with integer ids only). Do not include the locus name in the identifier in the FASTA file. Please note that you can not use this page to upload sequences for loci with extended attributes.	
Enter parameters	
locus:! abcZ	
status: Sanger trace checked	
sender:1 Jolley, Keith (keith) sequence isolate1 solate1 TTGATACCGTTGCCGAAGGTTCCGGCGAAATTCGTGATTTATTGCGCCGCTATCATCAT (FASTA):1 TTTGATACCGTTGGAAAATGGTTCGAGCGGCTTGTGTGAAAGAACTCAACCGAACTGG GGGGAACTCGGCTTGGCGGAACTGGCGGCATGTGCGGAACTGGCGGCATCAGCGAACTG GGGGAACTCGGCTTGGCTGCGGAACTGGGCGCTTTGCTGGAAACTGGCCGCGTCAGCCGGCG GGGGACTCGGTTGGCTGAGTGGGGTGCGAAAATGGCCGACGTTTGCGGAAATGGCGCGGTGACCTGGTGGTGGTGGAACTGGGCGGTAGCGGCGGTGAAACGGCGGTGGCCGGTGGTCGGGGTGGCGGCGGCGGCGGCGG	
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	
Use next available id (only for loci with integer ids)	
Action	
RESET	

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'.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Batch insert sequences	
Batch insert sequences	🛓 Help 🕜 🚯
Sequence check Locus: abcZ Original designation Allele id Status isolate1 1079 OK isolate2 1080 OK UPLOAD VALID SEQUENCES	

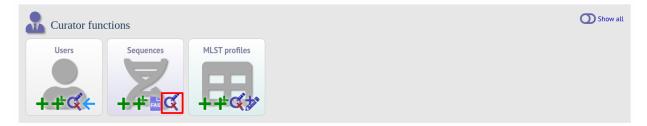
Any invalid sequences will be indicated in this confirmation page and these will not be uploaded (you can still upload the others), e.g.

PubMLST				
Home > Organisms >	• Organism > Neisseria ty	ping > Batch insert sequences		
Batch inse	ert sequences			💄 Help 🗹 🚺
Sequence che Locus: abcZ	eck			
Original designatio	n Allele id	Status	Action	
isolate1	1079	OK		
isolate2	1080 Sequence cont	ains non nucleotide (A C G T) characters.	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 update/delete sequences link.



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 (click the small arrow next to 'Filter query by' to expand the filter) and enter the allele number in the allele_id field.

PubMLST
Home > Organisms > Organism > Neisseria typing > Sequence attribute search
Sequence attribute search
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).
Search criteria Display
allele id v = 4 Order by: locus v ascending v Display: 25 v records per page ()
locus: abcZ status: type allele:
sender: Curator:
allele flag:

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.

PubMLST
Home > Organisms > Organism > Neisseria typing > Sequence attribute search
Sequence attribute search
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).
Search criteria Display
allele id v = v1 Order by: locus v ascending v Display: 25 v records per page (j)
- Filter query by Action
locus: abcZ v C RESET SUBMIT
type allele:
sender: curator: ()
allele flag:
1 record returned.
——Delete —— ——Database configuration — ——Flags ——
Delete ALL Export configuration/data Batch set
Delete Update locus allele sequence length status type allele sender curator date entered datestamp comments flags
AbcZ 1 TTTGATACTGTTGCC 433 Sanger trace Keith Jolley Keith Jolley 2001-02-07 2001-02-07 TTGTCGAACTCGATC Checked

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. You can also choose to delete and retire the allele identifier. If you do this, the allele identifier will not be re-used. It is possible to set the configuration so that you only have the option to delete and retire.

PubMLST		
Home > Organisms > Organis	m > Neisseria typing > Delete allele sequence	
Delete allele se	equence	2
You have chosen to delete th	e following record. Select 'Delete and Retire' to prevent the identifier being reused.	
locus: allele id: sequence:		Z,
status:	Sanger trace checked	
type allele:		
	Keith Jolley	
curator: date entered:	Keith Jolley	
	2001-02-07	
comments:		
Action DELETE AND RET	IRE	

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.

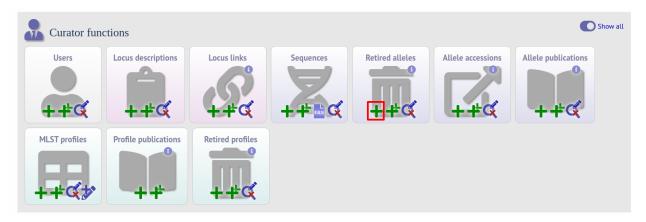
Note: Adding flags and comments to an allele record requires that this feature is enabled in the *database configuration*.

PubMLST		
Home > Organisms > Organism > Neisseria typing > Update allele sequence		
Update allele sequence		20
Please fill in the fields below - required fields are marked with an exclamation mark (!).		
Record	Action	
locus: l abcZ allele id: l sequence: TTTGATACTG_TTGCCGAAGG_TTTGGGCGGAA ATTCGCGATT_TATTGCGCGCG_TTATCATCAT GTCAGCCATC_AGTTGGAAAA_TGGTTCGAGT GAGGCCCTTAT_TGAAAGAGC_TCAACGAATTG GAACCTTGAGA_TCGAAGGGGAA_GGACGGCTGG	RESET	Å
AGTTGGATG CGGCGGTGAA GCAGACTTTG status: Sanger trace checked sender: Solley, Keith (keith) curator: Keith Jolley (keith) date entered: 2001-02-07 datestamp: 2020-07-22 type allele: Urue false New allele searches can be constrained to use just type alleles in comparisons comments:		
Flags: alternative start codon atypical contains IS element downstream fusion frameshift	A	
PubMed ids:		
ENA ids:		
Genbank ids:		
Override sequence length check		

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. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new retired allele id	
Add new retired allele id	
Please fill in the fields below - required fields are marked with an exclamation mark (!).	
Record Action	
locust NEIS0844	
allele id: 167	
datestamp: 2020-07-22	

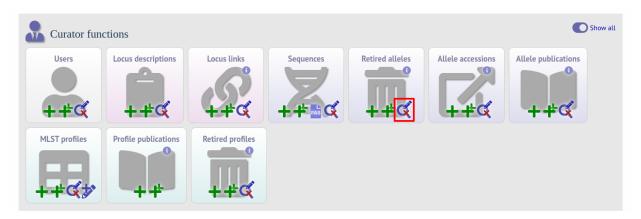
You cannot retire an allele that already exists, so you must delete it before retiring it. Once an identifier is retired, you will not be able to create a new allele with that name.

You can also retire an allele identifier when you delete an allele.

6.5 Un-retiring allele identifiers

If an allele identifier has been retired, it is possible to un-retire it so that it can be re-assigned. To do this, you need to remove the identifier from the retired_alleles table.

First, find the allele identifier in the retired_alleles table by clicking the 'Update/delete' retired alleles link on the sequence database curators' page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



Search by any criteria to find the allele identifier.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Query retired allele id information	
Query retired allele id information	
Please enter your search criteria below (or leave blank and submit to return all records).	
Display Display	
datestamp itoday itoday	
Filter query by Action	
locus: abcZ V D RESET SEARCH	
1 record returned.	
Delete	
Delete ALL	
Delete Update locus allele id curator datestamp × abcZ 1079 Keith Jolley 2020-08-05	

Click the delete link on the identifer to be un-retired.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Query retired allele id information	
Query retired allele id information	
Please enter your search criteria below (or leave blank and submit to return all records).	
Search criteria Display	
datestamp • = • today + Order by: locus • ascending • Display: 25 • records per page ()	
Filter query by Action	
locus: abcZ curator: SEARCH	
1 record returned.	
Delete	
Delete ALL	
Delete Update locus allele id curator datestamp X abcZ 1079 Keith Jolley 2020-08-05	

A confirmation page will be displayed. Click 'Delete' to remove the identifier from the retired alleles table.



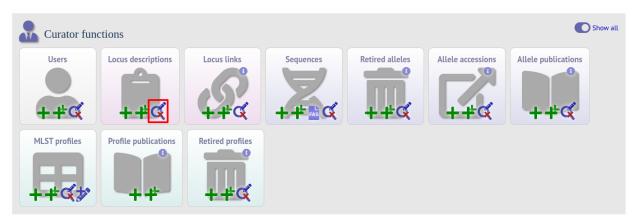
The identifier can now be re-assigned when adding a new allele.

6.6 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 in the 'locus descriptions' box. 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 in the 'locus descriptions' box 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). This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



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

PubMLST		
Home > Organisms > Organism > Neisseria typing > Query locus description in	formation	
Query locus description information		20
Please enter your search criteria below (or leave blank and submit to return all re	ecords).	
Search criteria	Display	
locus v = v NEIS0620 + (Order by: locus v ascending v	
	Display: 25 v records per page 🗊	
C Filter query by Action		
RESET		

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Query locus description information	
Query locus description information	
Please enter your search criteria below (or leave blank and submit to return all records).	

cus
Display: 25 ▼ records per page ()
locus: NEISO620 (maeA)
curator:
common name:

Click 'Submit'.

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Query locus description information	
Query locus description information	
Please enter your search criteria below (or leave blank and submit to return all records).	
Search criteria Display	
locus • • Order by: locus • ascending • Display: 25 • records per page ()	
locus: NEISO620 (maeA)	
curator: Common name: Common	
1 record returned.	
Delete Database configuration	
Delete ALL Export configuration/data	
Delete Update Locus full name product description curator datestamp X V NEIS0620 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:

PubMLST			
Home > Organisms > Organism	m > Neisseria typing > Update locus description		
Update locus d	lescription		2 0
Record locus:	- required fields are marked with an exclamation mark (!).	Action RESET SUBMIT	
datestamp:!	2020-07-22		
full name:			
product:	malate oxidoreductase (EC 1.1.1.38)		
description:	Final step in TCA cycle producing oxaloacetate.		
aliases:	NG09240 ÷		
PubMed ids:	14917678		
links: (Format: URL description)	<u>http://www.enzyme</u> -database.org/ <u>query.php</u> ? 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

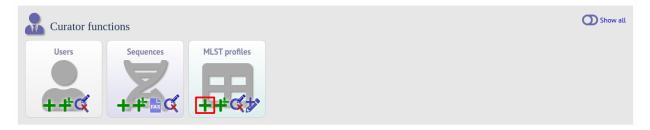
Enter links to additional web-based resources. Enter the URL first followed by a pipe symbol (I) and then the description.

Click 'Submit' when finished.

6.7 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 in the box named after the scheme name (e.g. MLST):



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	
Home > Organisms > Organism > Neisseria typing > Add new profile	
Add new MLST profile	💄 Help 🔀 🚯
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record ST: ! 15581 abc2: ! 2 adk: ! 3 aroE: ! 4 fumC: ! 122 gdh: ! 8 pdhC: ! 4 pgm: ! 6 sender: ! Jolley, Keith (keith)	ľ
clonal complex:	
curator: ! Keith Jolley (keith) date_entered: ! 2020-07-22 datestamp: ! 2020-07-22 PubMed ids:	
Action SUBMIT	

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.

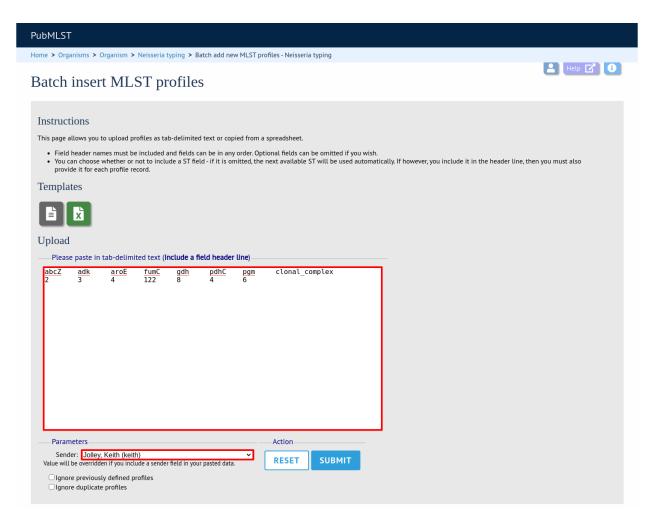
Click the batch add (++) profiles link next to the scheme name:

Curator fun	octions		O Show all
Users + + + C	Sequences + + Fas C	MLST profiles	

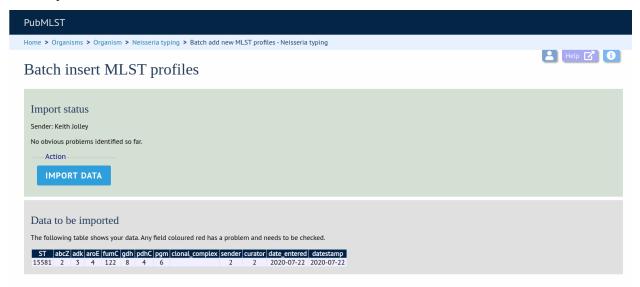
Click the 'Download submission template (xlsx format)' link to download an Excel submission template.

PubMLST
Home > Organisms > Organism > Neisseria typing > Batch add new MLST profiles - Neisseria typing
Batch insert MLST profiles
Instructions The page allows you to upload profiles as sub-delimited text or copied from a spreadsheet. a. i.e. i.e. i.e. i.e. i.e. i.e. i.e. i.
Parameters Action Sender: Select sender Value will be overridden if you include a sender field in your pasted data. RESET Ignore previously defined profiles Ignore duplicate profiles

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'.



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



6.8 Updating and deleting scheme profile definitions

In order to update or delete a scheme profile, first you must select it. Click the update/delete profiles link in the scheme profiles box named after the scheme (e.g. MLST):

Curator fund	ctions		Show all
Users	Sequences	MLST profiles	

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Query or update profiles	
Query or update profiles	🛓 Help 📝 🕄 🗡
Schemes Please select the scheme you would like to query: MLST Select	
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values. Locus/scheme fields ST V 4563 Order by: ST V ascending Display: ST V ascending V Display: ST V ST V	Action RESET SUBMIT
1 record returned. — Delete —— Delete ALL	
Detete Ubdate ST abcZ adk aroE fumC gdh pdhC pgm clonal complex 4563 2 7 6 13 9 18 8 ST-167 complex Analysis tools	
Export: Profiles Sequences	

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'. You can also choose to delete and retire the profile identifier. If you do this, the profile identifier will not be re-used. The database configuration can be set so that you can only delete and retire.

PubMLST		
Home > Organisms > Organis	m > Neisseria typing > Delete profile	
Delete profile		2
You have chosen to delete th	e following record. Select 'Delete and Retire' to prevent the identifier being reused.	
abcZ : adk : aroE : fumC : gdh : pdhC : pgm : clonal_complex : sender: curator: date entered:	4563 2 7 6 13 9 18 8 8 5T-167 complex Ana-Belen Ibarz-Pavon Keith Jolley	
Action		
DELETE DELE	TE AND RETIRE	

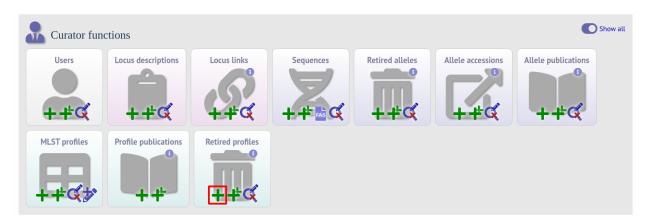
To modify the profile, click the 'Update' link next to the profile following the query. A form will be displayed - make any changes and then click 'Update'.

PubMLST		
Home > Organisms > Organism > Neisseria typing > Update profile		
Update profile		🛓 Help 🗹 🚯
Record	Action	
Update your record as required - required fields are marked with an exclamation mark (!): ST: ! 4563 abc2: ! 2 adk: ! 7 aroE:! 6 fumC:! 13 gdh: ! 9 pdhC: ! 18 pgm: ! 8 clonal_complex: ST-167 complex	RESET SUBMIT	
sender: ! Ibarz-Pavon, Ana-Belen (aibarz) ~ curator: ! Keith Jolley (keith)		
date_entered: ! 2005-03-03		
datestamp: ! 2020-07-22		
PubMed ids:		

6.9 Retiring scheme profile identifiers

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

You can retire a profile identifier by clicking the 'Add' link in the 'Retired profiles' box on the sequence database curators' page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



Select the scheme from the dropdown list box and enter the profile id. Click 'Submit'.

PubMLST		
Home > Organisms > Organism > Neisseria typing > Add new retired profile		
Add new retired profile		
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record scheme id:! MLST (id 1) profile id:! [7/232] curator:! Keith Jolley (keith) datestamp:! 2020-07-22	Action RESET SUBMIT	ľ

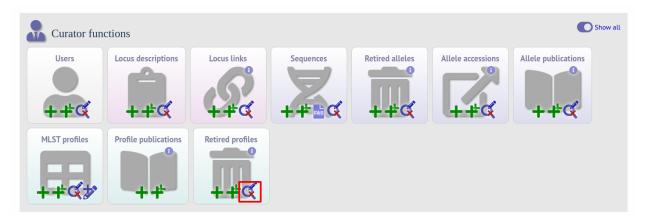
You cannot retire a profile identifier that already exists, so you must delete it before retiring it. Once an identifier is retired, you will not be able to create a new profile with that name.

You can also retire a profile definition when you *delete a profile*.

6.10 Un-retiring scheme profile identifiers

If a profile identifier, e.g. ST, has been retired, it is possible to un-retire it so that it can be re-assigned. To do this, you need to remove the identifier from the retired_profiles table.

First, find the profile identifier in the retired_profiles table by clicking the 'Update/delete' retired profiles link on the sequence database curators' page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



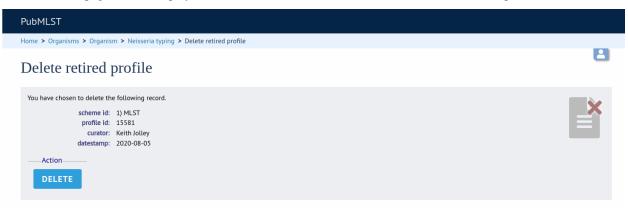
Search by any criteria to find the profile identifier.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Query retired profile information	
Query retired profile information	
Please enter your search criteria below (or leave blank and submit to return all records).	
Search criteria Display	
datestamp itoday <lit< td=""><td></td></lit<>	
Filter query by Action	
curator:	
1 record returned.	
Delete	
Delete ALL	
Delete Update scheme id profile id curator datestamp X Image: MLST (id 1) 15581 Keith Jolley 2020-08-05	

Click the delete link on the identifer to be un-retired.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Query retired profile information	
Query retired profile information	2 0
Please enter your search criteria below (or leave blank and submit to return all records).	
Search criteria Display	
datestamp • = • today + () Order by: scheme id • ascending • Display: 25 • records per page ()	
Filter query by Action	
scheme: MLST Curator: SEARCH	
1 record returned.	
Delete	
Delete ALL	
Delete Update scheme id profile id curator datestamp X MLST (id 1) 15581 Keith Jolley 2020-08-05	

A confirmation page will be displayed. Click 'Delete' to remove the identifier from the retired profiles table.



The identifier can now be re-assigned when adding a new profile.

6.11 Adding isolate records

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



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	
Home > Organisms > Neisseria spp. > Neisseria isolates > Add new isolate	
Add new isolate	💄 Help 🚺 🕄
Please fill in the fields below - required fields are marked with an exclamation mark (!).	
Primary metadata	
id:! 81776	
isolate: J323_2	
country: UK 🗸 🗸 🕥	
species:! Neisseria meningitidis	
sender: I Jolley, Keith (keith)	
curator:! Keith Jolley (keith) 🕧	
date entered: 2020-07-23 🕦	
datestamp: 2020-07-23 ①	
region: Enter one per line	
Supports multiple values - enter each one on separate line 🕢	
year: 2014 🚯	
date sampled: dd/mm/yyyy 🗖 🕦	
date received: dd/mm/yyyy 🗖 🕦	
non culture: O true O false 🕦	
epidemiological year:	
age yr:	
age range:	
age mth:	
sex: 🗙 🔿	

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.

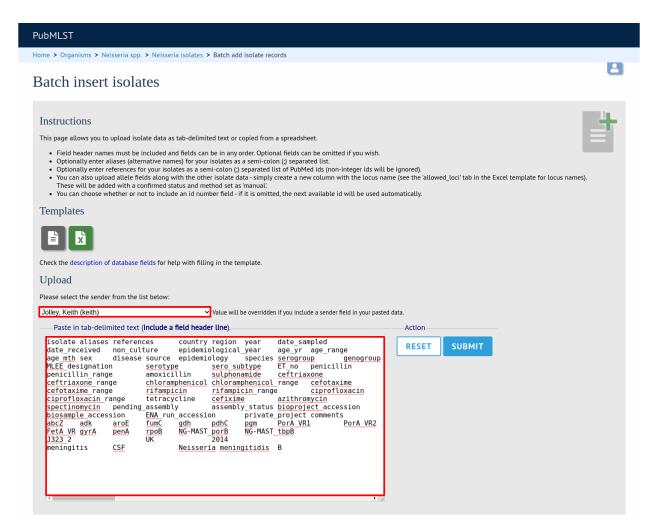
Curator fund	ctions			O Show all
Users	Isolates +++	Sequence bin	Sequence tags	

Download a submission template in Excel format from the link.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add isolate records	
Batch insert isolates	8
Instructions This page allows you to upload isolate 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. • Optionally enter aliases (alternative names) for your isolates as a semi-colon (1) separated list. • Optionally enter references for your isolates as a semi-colon (2) separated list of PubMed ids (non-integer ids will be ignored). • 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 the three will be added with a confirmed status and method set as 'manual'. • You can choose whether or not to include an id number field - if it is omitted, the next available id will be used automatically. Templates Image: Image: Ima	emplate for locus names).
Check the description of database fields for help with filling in the template.	
Upload	
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).	

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).



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

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add isolate records	
Batch insert isolates	2
Import status Sender: Keith Jolley No obvious problems identified so far. Action	
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 laliases references country region year date_sampled date_received non_culture epidemiological_year age_yr age_range age_mth sex disease source epidemiology 81776 J323_2 UK 2014	y speci Neisse mening

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

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add isolate records	
Batch insert isolates	2
Primary key Problem(s) id: 81776 species value 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. year date_sampled_date_received_non_culture_epidemiologicat_year_age_yr_age_range_age_mth_sexdiseasesource_epidemiology_speciesserogroup_genogroup_notes 2014 meningitis_CSFNeisseriaB	MLE
4 mengitidis	•

6.12 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.



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.

PubMLST											
Home > Organisms > Neisseria spp. > N	eisseria isolates > Qu	ery or update isolat	es								
Query or update isc	lates							E	Help	2](
Enter search criteria or leave blank to b	owse all records. Modi	fy form parameters	to filter	or enter a list of value	5.						
				Display/sort optic	ns						
Combine with: AND 🗸				Order by: id				✓ asce	ending 🗸		
country 👻 =	✓ USA	+	i	Display: 25	 records per page 	• (i)					
year 👻 =	✓ 2009			Action							
154 records returned (1 - 25 displayed).				RESET	ARCH						
— Delete — Tag scanning —	Projects				-Bookmark quer	v					
Delete ALL Scan	Select project			✓ Link 2	020-07-23:1	A	dd bookmar	k			
Conjunta Naji	6 () (>)		Isolate f	ields (j)			м	LST	Fine	typing anti	gens
	id isolate a	aliases countr	y year	disease	species	capsule group	ST	clonal complex	PorA_VR1	PorA_VR2	FetA_VR
× 🖋 土 🕂 12	674 M18700	USA	2009	invasive	Neisseria	В	7301	ST-32	+	+	+
× 🖋 土 + 12	675 M18701	USA	2009	(unspecified/other) invasive (unspecified/other)	meningitidis Neisseria meningitidis	В	7302	complex	+	+	+
	676 M18725		2009	invasive (unspecified/other)	Neisseria meningitidis	В	7303	ST-1572 complex	+	+	+
× 🖋 主 + 13	090 M19024 PA	A09015 USA	2009	meningitis	Neisseria meningitidis	В	7575		+	+	+

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

Pressing 'Delete' from this record page confirms the deletion. You can also choose to delete and retire the isolate. If you do this, the isolate id number will not be re-used. It is possible to set the configuration so that you only have the option to delete and retire.

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

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Update isolate	
Update isolate	💄 [Help 🔽 🚺
Primary metadata	
id: 12676 (i)	
isolate: M18725	
country:! USA 🔹 🗸	
species:! Neisseria meningitidis	
sender:! Zhao, Xin (xinzhao)	
curator:! Keith Jolley (keith)	
date entered: 2009-03-24 (j)	
datestamp: 2020-07-23 👔	
region: WI	
Supports multiple values - enter each one on separate line 👔	
year: 2009 🕦	
date sampled: dd/mm/yyyy 📋 👔	
date received: dd/mm/yyyy 🗂 🕦	
non culture: O true O false 👔	
epidemiological year:	
age yr:	
age range:	
age mth:	
sex:	
disease: invasive (unspecified/other)	

Allele designations can also be updated by clicking within the scheme tree and selecting the 'Add' or 'Update' link next to a displayed locus.

persons kissed past week: V () regular partner: O true O false () regular partner smokes: O true O false () ethnicity: V
regular partner smokes: Otrue Ofalse 🗿
ethnicity: 🖌 🗸
Loci
TypingMLST
All C adk aroE fumC gdh pdhC pgm ST clonal complex = @HLST (20 locus partial genes)
Locus: 16S_rDNA V Add/update

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'.

pubs clubs per week: persons kissed past week: regular partner: regular partner smokes: ethnicity:	▼ 0 ○ true ○ false ① 0 ○ true ○ false ① 0
Loci All loci Diping Locus: abcZ	Navigate and select schemes within tree to display allele designations

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.

6.13 Batch updating multiple isolate records

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

Curator fund	ctions			Show all
Users	Isolates + + +	Sequence bin	Sequence tags	

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'.

PubMLST		
Home > Organisms > Neisseria spp. > Neisseria isolates	> Batch Isolate Update	
Batch isolate update		Help 🔽
secondary fields are unique), this should be entere	ignored. e field that you are selecting isolates on). If a secondary selection	
2 abcZ 5 • The columns should be separated by tabs. Any oth • If you wish to blank a field, enter ' blank>' as the v Please enter the field(s) that you are selecting isolates o id will be used.	alue. n. Values used must be unique within this field or combination of	fields, i.e. only one isolate has the value(s) used. Usually the database
Please paste in your data below: id field value 100 serogroup B 101 serogroup B	Options Primary selection field: id Optional selection field: <none></none>	Allele designations Oreplace existing designation
	Cupdate existing values	Multi-value fields • Add additional value OReplace existing values

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.

PubMLST								
Home > Organisms	> Neisseria	a spp. > Ne	eisseria isolates > Batch Isolate	e Update				
Batch iso	late u	pdate	2					💄 Help 📝
The following cha	-		the database. Please check that t		nen press 'Update'. If	you do not wish to mał	ke these changes, press y	your browser's back button.
	serogroup			update field with new value				
2 101	serogroup	В	C	update field with new value				
Action UPDATE	Ī							

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
CN103 UK serogroup B	CN100	UK	serogroup	В
	CN103	UK	serogroup	В

6.14 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 up-date/delete isolates link:



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

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Query or update isolates	
Query or update isolates	💄 Help 🗹 🚯 🗡
Enter search criteria or leave blank to browse all records. Modify form parameters to filter	r or enter a list of values.
Isolate provenance fields	Display/sort options
Combine with: AND v date entered v = v 2014-03-18 + ()	Order by: id vascending v Display: 25 v records per page (
curator (surname) v = v Jolley	Action
	RESET SEARCH
3 records returned.	
——Delete —— Tag scanning — Projects —————	Bookmark query
Delete ALL Scan Select project	✓ Link 2020-07-23:1 Add bookmark
Delete Update Sequence bin New version	
id isolate aliases country year disease	species capsule group PorA_VR1 PorA_VR2 FetA_VR Neisseria meningitidis 5-1 2-81 2-81
🗙 🥜 主 🕂 28788 M22553 USA	Neisseria meningitidis 5-2 🖉 10-96 🖉 🕂
🗙 🥒 🏦 🕂 28789 M22568 USA	Neisseria meningitidis 7-2 🥒 4-39 🇨 🕇

You will have a final chance to change your mind:

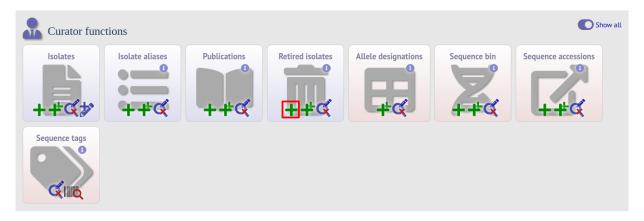
PubMLST			
Home > Organism	s > Neisseria spp. > Neisseria isolates > Delete multiple records		
Delete m	ultiple isolate records		
Warning		Confirm action	
	If you proceed, you will delete 3 isolate records. Please confirm that this is your intention.	DELETE DELETE AND RETIRE	
	The identifiers will not be re-assigned if you 'delete and retire'.		

Click 'Delete'. You can also choose to delete and retire the isolate id. If you do this, the id number will not be re-used. It is possible to set the configuration so that you only have the option to delete and retire.

6.15 Retiring isolate identifiers

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

You can retire an isolate identifier by clicking the 'Add' retired isolates link on the isolates database curators' page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



Enter the isolate id to retire and click 'Submit'.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Add new retired isolate i	1
Add new retired isolate id	
Please fill in the fields below - required fields are marked with an exclamation mark in Record Action isolate id: 46262 curator:! Keith Jolley (keith) datestamp:! 2020-07-23	

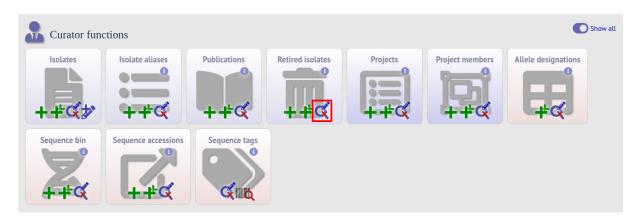
You cannot retire an isolate identifier that already exists, so you must delete it before retiring it. Once an identifier is retired, you will not be able to create a new isolate record using that identifier.

You can also retire an isolate identifier when you delete an isolate record.

6.16 Un-retiring isolate identifiers

If an isolate identifier has been retired, it is possible to un-retire it so that it can be re-assigned. To do this, you need to remove the identifier from the retired_isolates table.

First, find the isolate identifier in the retired_isolates table by clicking the 'Update/delete' retired isolates link on the isolate database curators' page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



Search by any criteria to find the isolate identifier.

PubMLST		
Home > Organisms > Neisseria spp. > Neisseria isolates >	Query retired isolate id information	
Query retired isolate id infor	mation	
Please enter your search criteria below (or leave blank and s	submit to return all records).	
Search criteria	Display	
datestamp v = v today	+ () Order by: isolate id v ascending v	
	Display: 25 v records per page (j	
Filter query by	Action	
curator: Jolley, Keith (keith)	CESET SEARCH	
1 record returned.		
DeleteUpdateisolate idcuratordatestampXImage: State of the state of t		

Click the delete link on the identifer to be un-retired.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Query retired isolate id information	
Query retired isolate id information	
Please enter your search criteria below (or leave blank and submit to return all records).	
Search criteria Display	
datestamp v = v today + isolate id v ascending v Display: 25 v records per page (j	
Filter query by Action	
1 record returned. Delete Update isolate id curator datestamp X Ø0516 Keith Jolley 2020-08-05	

A confirmation page will be displayed. Click 'Delete' to remove the identifier from the retired isolates table.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Delete retired isolate id	
Delete retired isolate id	
You have chosen to delete the following record. isolate id: 80316 curator: Keith Jolley datestamp: 2020-08-05 Action DELETE	

The identifier can now be re-assigned when adding a new isolate record.

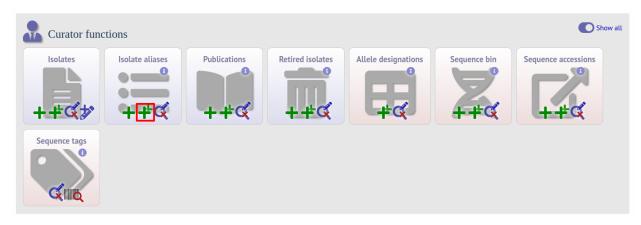
6.17 Setting alternative names for isolates (aliases)

Isolates can have any number of alternative names that they are known by. These isolate aliases can be set when isolates are first added to the database or batch uploaded later. When querying by isolate names, the aliases are also searched automatically.

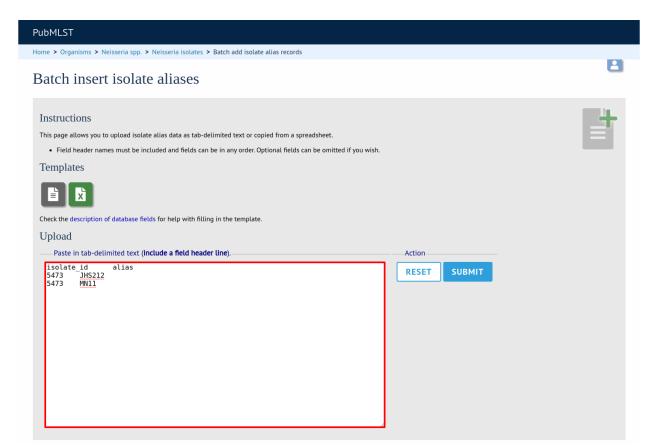
If adding isolates singly, add the aliases in to the aliases box (one alias per line):

If batch adding isolates, they can be entered as a semi-colon (;) separated list in the aliases column.

As stated above, aliases can also be batch added. To do this, click the batch add (++) isolate aliases link on the curator's index page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



Prepare a list in a spreadsheet using the provided template. This consists of two columns: isolate_id and alias. For example, to add the aliases 'JHS212' and 'NM11' to isolate id 5473, the values to paste in look like:



A confirmation page will be displayed.

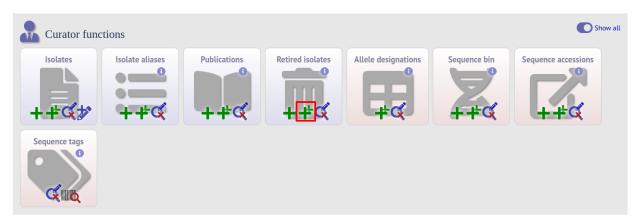
PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add isolate alias records	
Batch insert isolate aliases	2
Import status No obvious problems identified so far. Action IMPORT DATA	
Data to be imported The following table shows your data. Any field with red text has a problem and needs to be checked. Isolate_id_allasdatestamp_curator 5473_HK512_2020-07-232 5473_MN11_2020-07-232	

Click 'Import data'.

6.18 Linking isolate records to publications

Isolates can be associated with publications by adding PubMed id(s) to the record. This can be done when *adding the isolate*, where lists of PubMed ids can be entered in to the web form.

They can also be associated in batch after the upload of isolate records. Click the PubMed batch add (++) link on the curator's main page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



Open the Excel template by clicking the link.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add PubMed link records	
Batch insert refs	8
Instructions This page allows you to upload PubMed link 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. Templates Image: Im	Ē
Paste in tab-delimited text (include a field header line).	Action
	RESET

The Excel template has two columns, isolate_id and pubmed_id. Simply fill this in with a line for each record and then paste the entire spreadsheet in to the web form and press submit.

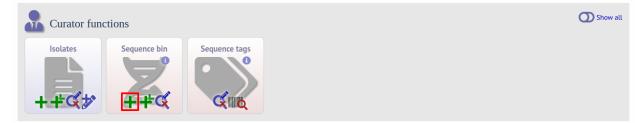
PubMLST		
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add PubMed link records		
Batch insert refs		
Instructions Inis page allows you to upload PubMed link data as tab-delimited text or copied from a spreadsheat. • Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish. Demplates Image: Im	Action RESET SUBMIT	

To ensure that publication information is stored locally and available for searching, the references database needs to be *updated regularly*.

6.19 Uploading sequence contigs linked to an isolate record

6.19.1 Select isolate from drop-down list

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



Select the isolate that you wish to link the sequence to from the dropdown list box (or if the database is large and there are too many isolates to list, enter the id in the text 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.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Add new sequences	
Upload sequences	
This page allows you to upload sequence data for a specified isolate record in FASTA format. 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:	¥
GGGTATGTAGTAATCAAATTTCCTGACTTATCTGTAAATACTTTAATTGTAGTTGTGGGT TGTCCACCTTCTTTAATAGAAGTAGTACCAATAACTTTTCCTAACTGCACAGTCCGCATA TATTGGCCATCAGGAGTCATCGGTAAGTGAAAAACTGAACGGTTATTCCTAATTGAGTATA ACCTTCAATTCATTTGGGAGATGGTAAAAACTGAACGGTTATTCCGCATTAGGCCTATGG AAGTGCCCCTCAAGAACATGCTCAAAGCCAGCAGATACCGGTTGTCCCGTTTGTCTCATT GGTGTATATCGAGACATGCTCATAGCCAGCGGATTCTGGGTTGCCCGTTTGTCTCATT TTTGATTCTTGTAGTAAGCCATCCAAGCCGAGTCCTCGGGTGGTATAGCCTGCTG GCCGCCTTCGGCAGCATCAAGCTCATCGGTTTTCGGGTTTGCCGCTTTGTCCGCTTG GCCGCCCTGGCGATCTCGCCCCGAAGGCGCTTGCCCAAGCGTGGTATAGCCTGCTG GCCGCCCTGGCGATCCCCGCCCCCCAAGGCGCTTCCCCAAGCCGTTGCCGCTACCGATCAGCCG GCCGCCCTGGCGATCCCCCCCCACAGGCGCTTGCCCCACGCATGACCGATACCCA GAAGGATTAGCCATGCAGGCGTCGTGATAGTTTTACCAGCCGGCCG	
Attributes Options isolate id: ! S sender: ! Jolley, Keith (keith) method: Illumina run id: Ill assembly id: Illumina Select FASTA file: Choose file Choose file No file chosen Action RESET	

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

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Add new sequences	

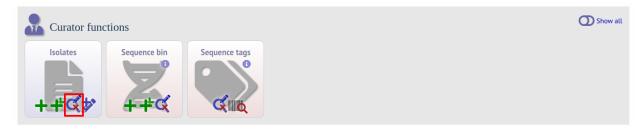
Upload sequences

he following sequ	ences will be e	entered.	Summary		Action
ginal designation S	equence length	Comments			
180426	26167	NODE_1211_length_26135_cov_16.536217	Number of contigs:		UPLOAD
180427	1240	NODE_619_length_1208_cov_9.455298	Minimum length:	265	
180428	1566	NODE_655_length_1534_cov_15.418513	Maximum length:	33,563	
180429	1580	NODE_675_length_1548_cov_17.753876	Total length:	2.059.411	
180430	8422	NODE_31_length_8390_cov_14.525030	Mean length:		
180431	2753	NODE_254_length_2721_cov_18.400587		,	
180432	1987	NODE_262_length_1955_cov_8.388747	N50 contig number:		
180433	30823	NODE_252_length_30791_cov_15.767627	N50 contig length (L50):	10,581	
180434	568	NODE_189_length_536_cov_31.078358	N90 contig number:	200	
180435	543	NODE_716_length_511_cov_12.113503	N90 contig length (L50):	3.126	
180436	13771	NODE_465_length_13739_cov_15.131669	N95 contig number:		
180437	15920	NODE_38_length_15888_cov_17.174660			
180438	2430	NODE_778_length_2398_cov_8.673060	N95 contig length (L50):	1,865	
180439	689	NODE_1765_length_657_cov_8.754947			
180440	287	NODE_729_length_255_cov_13.007843			
180441	16369	NODE_52_length_16337_cov_17.192997			
180442	684	NODE_190_length_652_cov_116.434052			
180443	3126	NODE_95_length_3094_cov_14.927279			
180444	5104	NODE_770_length_5072_cov_12.878943			
180445	1255	NODE_263_length_1223_cov_10.451349			
180446	4528	NODE_181_length_4496_cov_14.635231			
180447	4468	NODE_558_length_4436_cov_14.670198			
180448	949	NODE_1179_length_917_cov_13.504908			
180449	4065	NODE 527 length 4033 cov 15.608480			
180450	2467	NODE_212_length_2435_cov_12.223409			
180451	822	NODE_54_length_790_cov_9.398734			
180452	12793	NODE 146 length 12761 cov 14.578794			

6.19.2 Select from isolate query

As an alternative to selecting the isolate from a dropdown list (or entering the id on large databases), it is also possible to upload sequence data following an isolate query.

Click the isolate update/delete link from the curator's main page.



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	
Home > Organisms > Neisseria spp. > Neisseria isolates > Query or update isolates	
Query or update isolates	💄 Help 🗹 🚺 🥓
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.	
Isolate provenance fields Display/sort options	
isolate • = • M00242905 + () Display: 25 • rec	✓ ascending ✓ cords per page ()
Action SEARCH	
1 record returned.	
— Tag scanning— — Bookmark query—	
Scan 2020-07-23:1 Add bookmark	
Solate Index Sequence New Isolate fields ①	MLST Finetyping antigens
Delete Update Sequence New	species capsule ST clonal PorA_VR1 PorA_VR2 FetA_VR
	Neisseria B 1099 19 🖋 15 🖋 🕇

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.

Attributes isolate id: 1 3) M00242905 sender: 1 Jolley, Keith (keith) method: run id: assembly id:	Options Don't insert sequences shorter than 100 v bps. (j) Don't insert sequences containing only homopolymers. (j) Link to experiment: v Alternatively upload FASTA file Or enter Genbank accession Select FASTA file: Choose file No file chosen Action RESET SUBMIT

6.19.3 Upload options

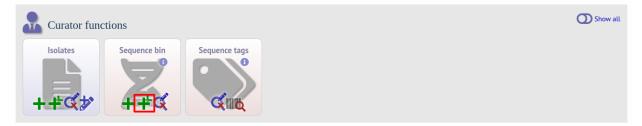
On the upload form, you can select to filter out short sequences or those containing only homopolymeric repeats (which can be artefactually produced by some assembler software versions) from your contig list.

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, select the experiment from the dropdown list box.

Attributes	Options
isolate id: 1 3) M00242905	☑ Don't insert sequences shorter than 100 v bps.
sender: ! Jolley, Keith (keith)	✓ Don't insert sequences containing only homopolymers. (i)
method:	Link to experiment:
run id:	Alternatively upload FASTA file or enter Genbank accession
assembly id:	
	Select FASTA file: Choose file No file chosen Action
	RESET

6.20 Batch uploading sequence contigs linked to multiple isolate records

To upload contigs for multiple isolates, click the batch add (++) sequence bin link on the curator's main page.



The first step is to upload the name of the contig file that will be linked to each isolate record. This can be done by pasting two columns in tab-delimited text format (e.g. from a spreadsheet) - the first column contains the isolate identifier, the second contains the filename of the contigs file, which should be in FASTA format.

You can choose which field to use for identifying the isolates, e.g. id (database id) or isolate (name of isolate). The value provided for this field needs to uniquely identify the isolate in the database - please note that only id is guaranteed to be unique.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add new sequences	
Batch upload sequence assemblies to multiple isolate records	
This function allows you to upload assembly contig files for multiple records together.	- 17
The first step in the upload process is to state which assembly contig FASTA file should be linked to each isolate record. You can use any provenance metadata field that uniquely identifies an isolate.	オ
You can upload up to 100 genomes at a time.	
——Identifying field name———	
Field; isolate	
Filenames	
Paste in tab-delimited text, e.g. copied from a spreadsheet, consisting of two columns. The first column should be the value for the isolate identifier field (specified above), and the second should be the filename that you are going to upload. You need to ensure that you use the full filename, including any suffix such as .fas or .fasta, which may be hidden by your operating system.	
JB 21292 JB 21292. fasta JB 21293 JB 21293. fasta JB 21294 JB 21294. fasta	
Action RESET SUBMIT	

Click Submit. The system will check to make sure that the isolate records are uniquely identified (if not, you will see an error message informing you of this and you will need to use the database id as the identifier). You will then see a file upload form.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add new sequences	
Batch upload sequence assemblies to multiple isolate records	20
Please upload the assembly contig files for each isolate record.	
remove row id isolate contigs current sequence bin state (contigs filename upload status 80309 JB_21292 - JB_21292.fasta X 80310 JB_21293 - JB_21292.fasta X 80311 JB_21294 - JB_21293.fasta X Remove 3 JB_21294.fasta X	E
Contig assembly files	
Please upload contig assemblies with the filenames as you specified (indicated in the table). Individual filesize is limited to 64 MB. You can upload up to 64 MB in one go, although you c upload multiple times so that the total size of the upload can be larger.	an
Drop files here or click to upload.	

Drag and drop your FASTA format contig files in to the dotted drop area. Provided the filenames exactly match the filename you stated, these will be uploaded to a staging area.

Click 'Validate' to check that these files are valid FASTA format.

PubMLS	T ganisms ➤ Neisser	ia spp. > Ne	eisseria isolates	Batch add new	sequences													
Batch	upload s	seque	nce asse	mblies t	o mult	iple	e is	so	olat	te	rec	oro	ls					Ð
remove row Remove All files up	80309 JB_2129 80310 JB_2129 80311 JB_2129 80311 JB_2129 bloaded. The seque	current se contigs 2 - 3 - 4 -	quence bin state total size (bp) - - - -	filename JB_21292.fasta JB_21293.fasta JB_21294.fasta	upload status	re they	y can	ı be i	addeo	d to t	he da	tabase					E	

The files will be checked and a table will be displayed showing the total sequence size and number of contigs found. Select the data sender and, optionally the sequencing method from the dropdown lists. Then click 'Upload validated contigs'.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add new sequences	

Batch upload sequence assemblies to multiple isolate records

Validation								
id isolate ff 80309 JB_21292 JB_2 80310 JB_21293 JB_2 80311 JB_21294 JB_2 You can upload 3 recording You can upload 3 recording	21292.fasta 21293.fasta 21294.fasta	valid FASTA	contig: 364 359 199	s total size 2,069,108 2,059,411 2,057,385				
Please do not refresh t	the page after	r you click t	he uplo	ad button. Up	load may take a :	short	while.	
Attributes							Options	
send	der: ! Jolley,	Keith (keith	1)			~	☑Don't insert sequences shorter than 100 ∨ bps. 👔	
meth	thod: Illumina	a	~				Don't insert sequences containing only homopolymers. (i)	
Action								
UPLOAD VAL	IDATED C	ONTIGS						

You can also choose to filter out short contigs selecting the checkbox and choosing the minimum length from the dropdown box in the options settings. You can also choose to filter out sequences containing only homopolymeric runs which can be produced artefactually by some assembler versions.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add new sequences	
Batch upload sequence assemblies to multiple isolate records	20
Validation id isolate filename valid FASTA contigs total size 80309 JB. 21292 JB. 21292.fasta 364 2,069,108 700 2009 100 2009 100	
80310 JB_21293 JB_21293.fasta 🗸 359 2,059,411 80311 JB_21294 JB_21294.fasta 🖌 199 2,057,385	
You can upload 3 records.	
Please do not refresh the page after you click the upload button. Upload may take a short while.	
Attributes Options sender: ! Jolley, Keith (keith) method: Illumina Action Don't insert sequences shorter than 100 v bps. (i) Don't insert sequences containing only homopolymers. (i) UPLOAD VALIDATED CONTIGS	

A confirmation message will be displayed after clicking the Upload button.

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add new sequences	
Batch upload sequence assemblies to multiple isolate records	
3 sequence assemblies uploaded.	

6.21 Linking remote contigs to isolate records

If *remote contigs have been enabled*, isolates can be linked to contigs stored in an external BIGSdb database, rather than directly uploaded. These well then be loaded when needed, for example during scanning or data export. This will be marginally slower than hosting contigs within the same database, but minimises duplication of sequence data and associated storage. Contigs need to be accessible via the BIGSdb *RESTful API*.

Click the sequences link icon on the curator's main page.

Curator fund	ctions					O Show all
Users	Isolates	Sequence bin	Sequence tags			
+ ₽ ₡<	+≠⊄⊅	+# <mark>@</mark> ¢	C III			

Either select the isolate id from the dropdown list, or enter it manually (list is disabled if there are >1000 records in the database). Enter the URI for the RESTful API of the parent isolate record, e.g. http://rest.pubmlst.org/db/pubmlst_rmlst_isolates/933. This URI can require authentication if credentials have been *set up*.

Press submit.

PubMLST	
Home > Organisms > Organism > Leptospira isolates > Add remote contigs	
Add remote contigs	
This page allows you to link contigs stored in a remote BIGSdb database to an isolate record. Access to these contigs is via the BIGSdb RESTful API which must be running on the remote database.	
Valid URIs are in the form 'http://rest.pubmlst.org/db/{database_config}/isolates/(isolate_id)'.	
Enter details Action	
isolate id: 1 Insolate record URI: 1 http://rest.pubmlst.org/db/pubmlst_rmlst_isolates/isolates/933	

Summary information about the number of contigs and their total length will be downloaded from the remote isolate record. You will then be prompted to upload this information to the database, by clicking the 'Upload' button.

PubMLST	
Home > Organisms > Organism > Leptospira isolates > Add remote contigs	
Add remote contigs	2
Checking contigs Downloading isolate record done.	
Contigs: 25 Total length: 2,697,907 bp	
UPLOAD	

The contigs will be downloaded in bulk in order to determine their lengths. This information is stored within the local database as it is required for various outputs. Full metadata is not stored at this stage.

PubMLST	
Home > Organisms > Organism > Leptospira isolates > Add remote contigs	
Add remote contigs	2
25 remote contigs added.	
Please note that so far only links to the contig records have been added to the database. The contigs themselves have not been downloaded and checked. Processing records the length of each contig and stores a checksum within the database so that it is possible to tell if the sequence ever changes. You can either do this now or it can be performed offline by a scheduled task.	
Total contigs: 25	
Remote contigs: 25 (25 unprocessed) Total length: 2,697,907	
Action PROCESS CONTIGS NOW	

This is all that is required for the contigs to be used as normal. In order to get the full metadata about the contigs (sequencing platform used, sender and datestamp information), you can choose to process the contigs by clicking the 'Process contigs now' button. This will download each contig in turn, and store its provenance metadata locally.

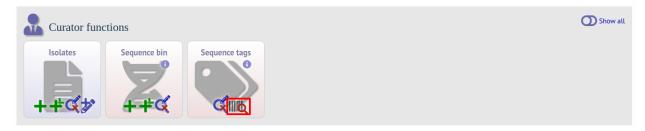
PubMLST	
Home > Organisms > Organism > Leptospira isolates > Add remote contigs	
Add remote contigs	2
Processed contigs Contigs processed: 25 Total length: 2,697,907 bp	

Alternatively, this step can be *performed offline automatically*.

6.22 Automated web-based sequence tagging

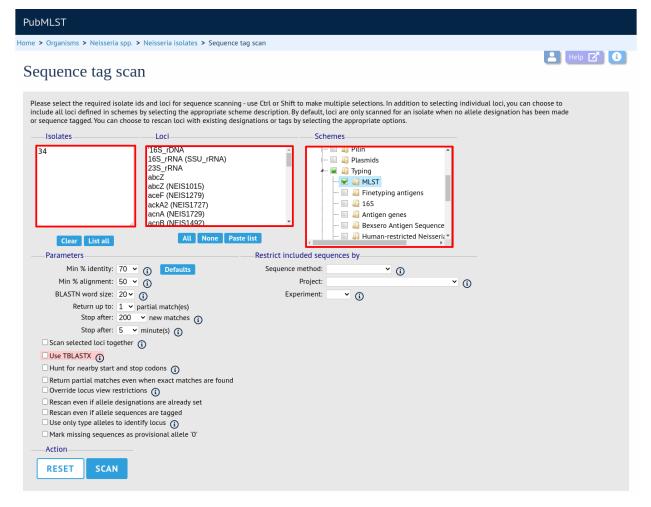
Sequence tagging, or tag-scanning, is the process of identifying alleles by scanning the sequence bin linked to an isolate record. Defined loci can either have a single reference sequence, that is defined in the locus table, or they can be linked to an external database that contains the sequences for 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.



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. On database with a large number of isolates, you will need to enter a list of isolate ids rather than pick from a 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

You can select to only use type alleles to identify the locus. This will constrain the search space so that allele definitions don't become more variable over time. If a partial match is found to a type allele then a full database lookup will be performed to identify any known alleles. An allele can be given a status of type allele when *defining*.

If fast scanning is enabled, there will also be an option to 'Scan selected loci together'. This can be significantly quicker than a locus-by-locus search against all alleles but is not enabled by default as it can use more memory on the server and requires *exemplar alleles* to be defined.

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

Parameters
Min % identity: 70 🖌 👔 🛛 Defaults
Min % alignment: 50 🖌 👔
BLASTN word size: 20 🗸 👔
Return up to: 1 🗸 partial match(es)
Stop after: 200 🗸 new matches 👔
Stop after: 5 🗸 minute(s) 👔
Scan selected loci together (i)
Use TBLASTX (
Hunt for nearby start and stop codons (
Return partial matches even when exact matches are found Override locus view restrictions (i)
Rescan even if allele designations are already set
Rescan even if allele sequences are tagged
Use only type alleles to identify locus (i)
Mark missing sequences as provisional allele '0'

Press 'Scan'. The system takes approximately 1-2 seconds to identify each sequence (depending on machine speed and size of definitions databases). Alternatively, if 'Scan selected loci together' is available and selected, it may take longer to return initial results but total time should be less (e.g. a 2000 loci cgMLST scheme may be returned in 1-2 minutes). Any identified sequences will be listed in a table, with checkboxes indicating whether allele sequences or sequence regions are to be tagged.

PubMLST

Home > Organisms > Neisseria spp. > Neisseria isolates > Sequence tag scan

Sequence tag scan

solate	Match	Locus	Allele	% identity	Alignment length	Allele length	E- value	Sequence bin id	Start	End	Predicted start	Predicted end	Orientation	Designate allele	Tag sequence	Flag 🧃
34) 20	exact	abcZ	1	100.00	433	433	0.0	182791	7064	7496	7064	7496 extract 🖸	+			
34) 20	exact	adk	3	100.00	465	465	0.0	182750	1392	1856	1392	1856 extract 🔊	→			
34) 20	exact	aroE	1	100.00	490	490	0.0	182731	32577	33066	32577	33066 extract 🖸	+			
34) 20	exact	fumC	1	100.00	465	465	0.0	182815	19783	20247	19783	20247 extract 🔊	→			
34) 20	exact	gdh	1	100.00	501	501	0.0	182852	7516	8016	7516	8016 extract 🔊	→			
34) 20	exact	pdhC	1	100.00	480	480	0.0	182819	13868	14347	13868	14347 extract 🔊	→			
34) 20	exact	pgm	3	100.00	450	450	0.0	182831	24559	25008	24559	25008 extract O	→			
Act TA		LELE	S/SEQ	DENCE	ES									All None	All None	
nished	Thu Jul I: Thu Ju time: 5	ul 23 1	5:37:03													

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

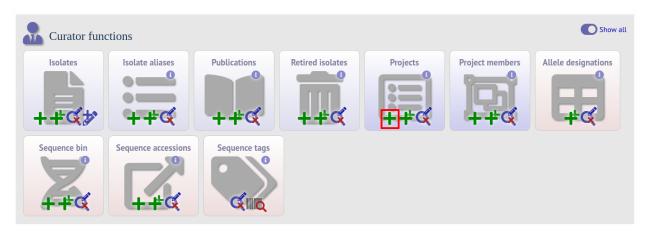
6.23 Projects

6.23.1 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. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.

💄 Help 📝 🚯



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.

There are a further two option flags:

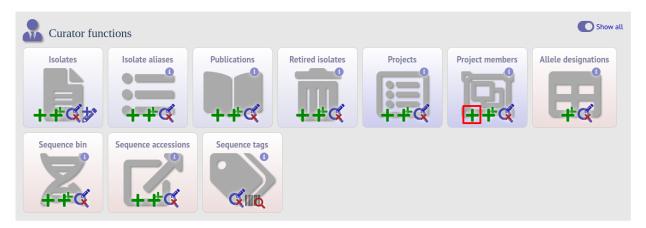
- private Setting this makes the project a private *user project*. You will be set as the project owner and will be the only user able to access it by default. You can add additional users or user groups who will be able to access and update the project data later.
- no_quota If set, isolates added to this project will not count against a user's quota of *private records* (only relevant to private projects).

Click 'Submit'.

PubMLST			
Home > Organisms > Neisseria spp. > Neisseria isolates > Add new project description			
Add new project description			2 8
Please fill in the fields below - required fields are marked with an exclamation mark (!).			
Record	Action		
id:! 3 short description:! MRF Meningococcus Genome Library	RESET	SUBMIT	
isolate display: O true 🖲 false 🎧			
list:! O true 🖲 false 👔			
private:! O true 🖲 false 🚺			
no quota:! 🖲 true 🔿 false 🐧			
restrict user:! O true 🖲 false 👔			
restrict usergroup: O true 🖲 false 👔			
curator: Keith Jolley (keith) datestamp: 2020-07-23			
full description: England, the Scottish Haemophilus, Legionella, Meningococcus and Pneumococcus Reference (<u>SHLMPR</u>) Laboratory, The <u>Wellcome</u> Trust Sanger Institute and the University of Oxford, funded by the Meningitis Research Foundation.			
curate config:			

6.23.2 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. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



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'.

PubMLST		
Home > Organisms > Neisseria spp. > Neisseria isolates > Add new project member		
Add new project member		
Please fill in the fields below - required fields are marked with an exclamation mark ((!).	
Record	Action	
project id:! 3) MRF Meningococcus Genome Library v isolate id:! 18968 curator:! Keith Jolley (keith) datestamp:! 2020-07-23	RESET SUBMIT	

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

Curator fund	ctions					Show all
Isolates	Isolate aliases	Publications	Retired isolates	Projects	Project members	Allele designations
Sequence bin	Sequence accessions	Sequence tags				

Download an Excel submission template:

PubMLST		
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add project member records		
Batch insert project members		2
Instructions 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. 		
Templates		
Check the description of database fields for help with filling in the template.		
Upload		
Paste in tab-delimited text (Include a field header line).	Action	
	RESET SUBMIT	

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'.

PubMLST		
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add project member records		
Batch insert project members		
Instructions 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. Templates Deck the description of database fields for help with filling in the template. Upload Paste in tab-delimited text (include a field header line). project_id isolate_id 3 18968	Action RESET SUBMIT	
3 18969 3 19023 3 19024 3 19026 3 19027 3 19028 3 19029 3 19031 3 19032 3 19033 3 19034		

See also:

Setting up user projects

User projects

6.24 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:

BIGSdb Home > Organisms > Neisseria s	spp. > Neisseria isolates > Isolate informat	tion			
Full information	on on isolate M1766	1 (id:193	77)		Help 🗹
Provenance/pri	imary metadata				
isolate: strain designation: country: continent: region: year: disease:	North America MI 2008 invasive (unspecified/other) Neisseria meningitidis W C	genogroup notes: capsule group:	C backbone: csc fragmented (71.17% cov),cssB fragmented (16.16% cov),cssB fragmented (84.4% cov),cttB fragmented (98.45% cov),cttF fragmented (92.3% cov),cttF fragmented (15.31% cov),cttF fragmented (81.43% cov),cttF fragmented (81.43% cov),missing cttD,missing tex. Prediction code: https://github.com/ntopaz/chara cterize_neisseria_capsule. discrepancy	curator: update history: date entered:	
Vaccines Bexsero reactivity:	insufficient data				
Versions More than one version of this Newer versions: Publication (1) • Kistyuk AO, Katz	40500	nan P, Nelakuditi V.	Humphrey JC, Sammons SA, Govil D, Maii	r RD, Tatti KM, Tondella	ML, Harcourt BH, Mayer LW, Jordan

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:



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

T GDT	1LST															
lome >	Organisr	ms > Neisse	eria spp. >	Nei	isseria isolate	s > Query or update	isolates									
Que	ery o	or upo	late i	SC	lates								E	Help		
Enter	search o	criteria or le	ave blank	to b	rowse all reco	rds. Modify form par	ameters to	filter or	enter a list of values.							
	Isolate j	provenance	e fields —						Display/sort option	IS						
ic	I		v =		✓ Ente	r value	+ (0	Order by: id					ascendi	ng 🗸	
									Display: 25 🗸	records per page	i					
/	Action-															
	RESET	SE/	ARCH													
		ls returned (1 - 25 disp	olaye	ed).											
— Ta	ag scani	ning — —	Projects	;					—Private records —	Bookmark o	query					
	Scan		Select pro	ject.			✓ Lin	k	Publish	2020-07-23:1		Ad	d bookmark			
«	Scan	1 2			6 🕟	\otimes	✓ Lin	k	Publish	2020-07-23:1		Ad	d bookmark			
«					\sim	>>				2020-07-23:1						
(«	•	1 2 Sequence	3 4 New	5	6 📀	>		k late field		2020-07-23:1	cancule		MLST	Finet	yping ant	Ť.
		1 2 Sequence bin	3 4 New version	5 id	6 🔊 isolate	aliases	lsol country	ate field	ds 👔 disease	species	capsule group			Finet PorA_VR1	PorA_VR2	2 FetA_VR
(«) Delete	•	1 2 Sequence	3 4 New	5	6 📀	aliases B1; NIBSC_2803; Z1001	Isol	ate fiel	ds (j)				MLST clonal	Finet		ī —
	•	1 2 Sequence bin	3 4 New version	5 id	6 🔊 isolate	B1; NIBSC_2803; Z1001 B35; NIBSC_2822;	lsol country	ate field	ds ① disease invasive (unspecified/other) meningitis and	species Neisseria meningitidis Neisseria	group		MLST clonal complex ST-1	Finet PorA_VR1	PorA_VR2	2 FetA_VR
×	•	1 2 Sequence bin	3 4 New version	5 id 1 2	6 > isolate A4/M1027	B1; NIBSC_2803; Z1001	lsol country USA	ate field year 1937	ds () disease invasive (unspecified/other) meningitis and septicaemia invasive	species Neisseria meningitidis Neisseria Neisseria	group A	ST	MLST clonal complex	Finet PorA_VR1 5-2 🖉	PorA_VR2	2 FetA_VR F1-5 🖋
×	•	1 2 Sequence bin	3 4 New version	5 id 1 2	6 > isolate A4/M1027 120M	B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035	Isol country USA Pakistan	ate field year 1937 1967	disease invasive (unspecified/other) meningitis and septicaemia	species Neisseria meningitidis Neisseria meningitidis	group A A	ST 1	MLST clonal complex ST-1	Finet PorA_VR1 5-2 0 5-2 0	PorA_VR2 10 2 10 2	2 FetA_VR F1-5 2 F5-1 2
× × × ×	•	1 2 Sequence t t t t	3 4 New version + + +	5 id 1 2 3 4	6) isolate A4/M1027 120M M00242905 M1027	B1; NIBSC_2803; Z1001 B35; NIBSC_2822;	Isol country USA Pakistan UK USA	ate field year 1937 1967 2000 1937	ds () disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	group A A B A	1 1099 4	MLST clonal complex ST-1 complex ST-4 complex	Finet PorA_VR1 5-2 @ 5-2 @ 19 @ +	PorA_VR2 10 🖉 10 🖉 15 🍼 +	2 FetA_VR F1-5 ♪ F5-1 ♪ + +
××××	•	1 2 Sequence bin 1 1 1 1 1 1 1 1 1	3 4 New version + + +	5 id 1 2 3 4	6 isolate A4/M1027 120M M00242905	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076;	Isol country USA Pakistan UK	ate field year 1937 1967 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive	species Neisseria meningitidis Neisseria meningitidis Neisseria Neisseria	group A A B	ST 1 1099	MLST clonal complex ST-1 complex ST-4	Finet PorA_VR1 5-2 5-2 19 2	PorA_VR2 10 a 10 a 15 a	2 FetA_VR F1-5 2 F5-1 2 +
× × × ×	•	1 2 Sequence t t t t	3 4 New version + + +	5 id 1 2 3 4 5	6) isolate A4/M1027 120M M00242905 M1027	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076;	Isol country USA Pakistan UK USA	ate field year 1937 1967 2000 1937	ds () disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria	group A A B A	1 1099 4	MLST clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22	Finet PorA_VR1 5-2 @ 5-2 @ 19 @ +	PorA_VR2 10 🖉 10 🖉 15 🧭 +	2 FetA_VR F1-5 ♪ F5-1 ♪ + +
× × × × ×	•	1 2 Sequence bin 1 1 1 1 1 1 1 1 1	3 4 New version + + + + +	5 id 1 2 3 4 5	6 isolate A4/M1027 120M M00242905 M1027 M00240227	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760;	Isol country USA Pakistan UK USA UK	ate field year 1937 1967 2000 1937 2000	ds disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive	specles Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria	group A A B A B	ST 1 1099 4 1100	MLST clonal complex ST-1 complex ST-4 complex ST-5 complex ST-5 complex ST-5	Finet PorA_VR1 5-2 0 5-2 0 19 0 + 7 0	PorA_VR2 10 / 10 10 / 15 / 10 15 / 16 / 16	FetA_VR F1-5 F5-1 + + +
× × × × ×	•	1 2 Sequence t t t t t t t	3 4 New version + + + + + + +	5 id 1 2 3 4 5 6 7	6 Solate A4/M1027 120M M00242905 M1027 M00240227 M00282207	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043	Isol country USA Pakistan UK USA UK	ate field year 1937 1967 2000 1937 2000 2000	disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria	group A A B A B W	1 1099 4 1100 1101	MLST clonal complex ST-1 complex ST-32 complex ST-32 complex ST-22 complex ST-5 complex ST-15	Finet PorA_VR1 5-2 3 5-2 3 19 3 + 7 3 +	PorA_VR2 10 0 10 0 15 0 + 16 0 +	FetA_VR F1-5 * F5-1 * + + + + + +
× × × × ×	•	1 2 Sequence bin 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2	3 4 New version + + + + + + +	5 id 1 2 3 4 5 6 7	6 () isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760;	Isol country USA Pakistan UK USA UK UK Finland	ate field 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) invasive	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	group A A B A B W A	ST 1 1099 4 1100 1101 5	MLST clonal complex ST-1 complex ST-2 complex ST-22 complex ST-22 complex ST-25 complex	Finet PorA_VR1 5-2 0 19 0 + 7 0 20 0	PorA_VR2 10 0 10 0 15 0 + 16 0 + 9 0	FetA_VR F1-5 // F5-1 // + + + + F3-1 //
× × × × × × ×	•	1 2 Sequence bin 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	3 4 New version + + + + + + +	5 id 1 2 3 3 4 5 5 6 7 8	6 () isolate A4/M1027 120M M00242905 M1027 M00242207 7891 M00242007	B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043 B54; NIBSC_2760;	Isol country USA Pakistan UK USA UK UK Finland	ate field year 1937 2000 1937 2000 2000 1975 2000	ds 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)	specles Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	group A A B A B W A B B	ST 1 1099 4 1100 1101 5 1102	MLST clonal complex ST-1 complex ST-4 complex ST-32 complex ST-52 complex ST-52 complex ST-52 complex	Finet PorA_VR1 5-2 5-2 19 + 7 20 +	PorA_VR2 10 0 10 0 15 0 + 16 0 + 9 0	FetA_VR F1-5 // F5-1 // + + + + F3-1 //

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.

PubMLST					
Home > Organisms > Neisseria	a spp. > Neisseria isolates > Create new	isolate record version			
Create new isc	late record version				E Help C
designations will not. This fa	te a new version of the isolate record show cilitates storage of different versions of g ry will be reset for the new record.				
Enter new record id	Options	Action]		
id: 80312	Add new version to projects	CREATE			
Provenance/pr	im anu matadata				
	0				
strain designation:	7 7891 B54; NIBSC_2760; Z1054 A: P1.20,9: F3-1: ST-5 (cc5)	genogroup: genogroup notes:	A A backbone: All essential capsule genes intact and present. Prediction code: https://github.com/ntopaz/chara		Wendell Zollinger, Dept Bacterial Diseases, Walter Reed Army Institute of Research, Washington DC, USA Auto Tagger
isolate: aliases: strain designation: country: continent: year:	7 7891 B54; NIBSC_2760; Z1054 A: P1.20,9: F3-1: ST-5 (cc5) Finland Europe 1975 invasive (unspecified/other)		A backbone: All essential capsule genes intact and present. Prediction code: https://github.com/ntopaz/chara cterize_neisseria_capsule. A Subgroup III 4,21	curator: update history: date entered:	Bacterial Diseases,Walter Reed Army Institute of Research, Washington DC, USA Auto Tagger 120 updates show details

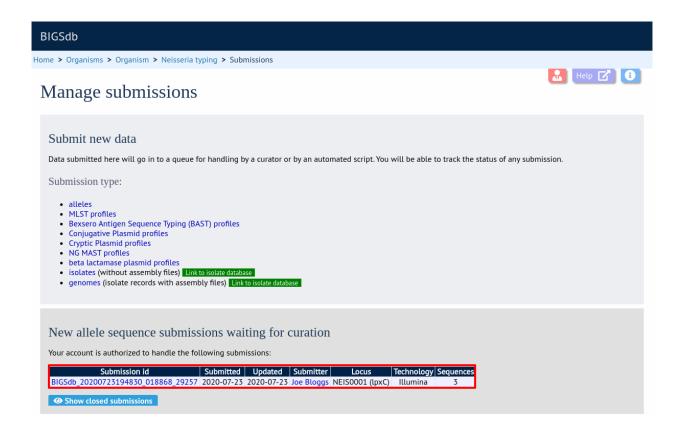
CHAPTER 7

Curating data submitted via the automated submission system

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.

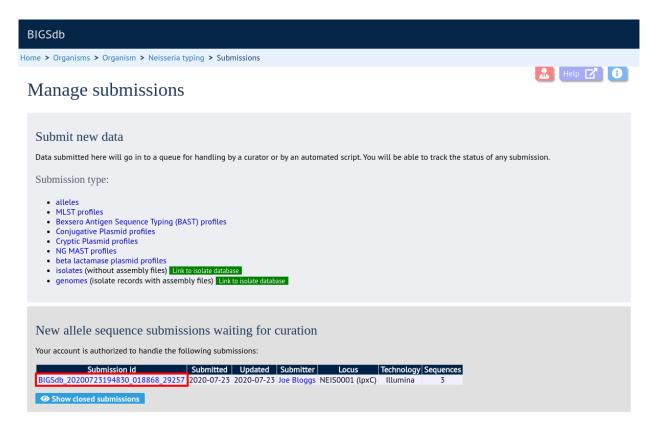
BIGSdb							
Home > Organisms > Orga	nism > Neisseria typin	g					
<i>Neisseria</i> typ	oing databa	se					ă.
		Query a	sequence			€ LOG OUT	+
	Sequence query		Batch sequence q	<u>uery</u>			
	Query a single seque genome assembly to matches.		Query multiple inde in FASTA format to i matches.				+
						EXPORT	+
		Find a	alleles			ANALYSIS	+
	Sequence attribut		Locus-specific sec search				+
	loci together)		Select, analyse and alleles from a single				+
		Search for a	llelic profiles			SISOLATES	
Allelic profile que	<u>ry</u>	Search by combin	nations of alleles	<u>Batch profile qu</u>	<u>iery</u>		
Search, browse or er	nter list of profiles	This can include pa related profiles.	rtial matches to find	Lookup multiple a together.	llelic profiles		

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



7.1 Alleles

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

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	🔝 Help 🗹 🚯
Submission: BIGSdb_20200723194830_018868_29257 Summary type: alleles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-23 status: pending locus: NEIS0001 (lpxC) sequences: 3 technology: Illumina read length: 100-199 coverage: 20-49x assembly: de novo assembly software: Spades	
Sequences Identifier Length Sequence Complete CDS Status Query Assigned allele UK322 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA	Messages Message: Append Send now Archive Archive of submission and any supporting files: Download

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

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	Help 🗹 🚯
Submission: BIGSdb_20200723194830_018868_29257	
type: alleles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-23 status: pending locus: NEISO001 (lpxC) sequences: 3 technology: Illumina read length: 100-199 coverage: 20-49x assembly: of envo assembly: addes a	
Sequences Identifier Length Sequence Complete CDS Status Query Assigned allele UK322 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGGGATAA pending < Q	Messages Message: Append Send now
UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA ✓ pending < Q	Archive Archive of submission and any supporting 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.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	n	
Curate submission		🔝 Help 🖍 🚯
Submission: BIGSdb_20200723194830_018868_29257		
type: alleles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-23 status: pending locus: NEIS0001 (lpxC) sequences: 3 technology: Illumina read length: 100-199 coverage: 20-49x assembly: de novo assembly: Spades		
Sequences		Messages
Identifier Sequence C UK322 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA UK323 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA BATCH CURATE BATCH QUERY	omplete CDS Status Query Assigned allele pending <	Message: Append Send now Archive Archive of submission and any supporting files: Download TAR

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		
Home > Organisms > Organis	sm > Neisseria typing > Add new allele sequence	
Add new allele	e sequence	Help 🗹 🗿
Please fill in the fields below	r-required fields are marked with an exclamation mark (!).	Action
locus:!	NEIS0001 V	
allele id: 1	1992 🔤	RESET SUBMIT
	ATCGATTTCGCCGGCAAATCCTACATCGACGAAATCGCGCCGCGCGCACTTTCGGCTTATGCACGAACTGGAA ATGATGCGCGCCCCACAATCTGGTTTGGCGCGCAATTGGACAACGCCATCGTGATTGACGACAGGATGCCTG ATGATGCTGAAGGCTTGCGCTATCCCGATGAGTTTGGCCGCCACAAATCCTTGATGCCATCGGCATTTGTATATC GTCGGACACCCGATTATCGGTGCGTTTGAAGGCTACAAATCGGGACACGCCATCAACAACGCACTCTTGCGCGG GTTTTGGCAGACGAACGGCTTACGACCGGGGGGGAATTTGCCCGACAGCGATGATTTGCCCCGACGACTTTCACGGG CTGAACATCAGAACTTGTGGATAA	
status:!	unchecked 🗸	
	Bloggs, Joe (jbloggs)	
	Keith Jolley (keith)	
date entered: 2 datestamp: 2		
	○ true ○ false New allele searches can be constrained to use just type alleles in comparisons	
comments:		
a C	alternative start codon atypical contains IS element	

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.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new allele sequence	
Add new allele sequence	Help 🕜 🕄
Sequence NEIS0001 (lpxC): 1993 added. Don't forget to close the submission!	

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.

Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission Curate submission Submission: BIGSdb_20200723194830_018868_29257
Curate submission Submission: BIGSdb_20200723194830_018868_29257
type: alleles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-23 status: pending locus: NEIS0001 (lpxC) sequences: 3 technology: Illumina read lengt: 100-199 coverage: 20-49% assembly: de novo assembly: de novo Archive Mark att: Pending: Rejected update
BATCH CURATE BATCH QUERY Archive of submission and any supporting files:

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.

PubMLST		
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission		
Curate submission		
Submission: BIGSdb_20200723194830_018868_29257		
Summary type: alleles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-23 status: pending locus: NEIS0001 (lpxC) sequences: 3 technology: Illumina read length: 100-199 coverage: 20-49x assembly: de novo assembly software: Spades		
Sequences		Messages
Identifier Length Sequence Co UK322 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTTGTGGATAA UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAACTTGTGGATAA BATCH CURATE BATCH QUERY	Implete CDS Status Query Assigned allele assigned pending Q Currate pending Q Currate Mark all: Pending Rejected Update 	Message: Append Send now Archive Archive of submission and any supporting files: Download

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.

ubMLST	
me > Organisms > Organism > Neisseria typing > Batch insert sequences	
Batch insert sequences	
This page allows 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 (loci with integer ids only). Do not include the locus name in the identifier in the FASTA file. Please note that you can not use this page to upload sequences for loci with extended attributes. Enter parameters	
locus: NEIS001 (lpxC)	
status: unchecked •	
sender:! Bloggs, Joe (jbloggs)	
sequence >UK323 ATGCTGCAAAGAACTTTGGCGAAATCCATCAGCGTTACCGGAGTCGGGCTGGATTCGGGGC GAACGCGTGCGCGCTGACCCTGCACCCGCGCCTGAAAACAGCTGGATTTCCTTCGGAAGT ACCGGCTGCTCCACCTTGCGACAAATCAAACTGACCCCTTATTTGATCAACGAT ACCGGCTTTCCCACCATCGTACCGACAAAGGCGTGCGGCACAATCGAACGAT ACCGGCTTGCCGCTTACCGACAAGGCGTGCGGCACGGCA	
🛿 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	
Use next available id (only for loci with integer ids)	
Action	
RESET CHECK	

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Batch insert sequences	
Batch insert sequences	🛓 Help 🚺 🕄
Sequence check Locus: NEIS0001 (IpxC) Original designation Allele id Status UK323 1992 OK UK347 1994 OK UPLOAD VALID SEQUENCES	

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Batch insert sequences	
Batch insert sequences	Leto C
Sequences added. Don't forget to close the submission!	

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

PubMLST
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission
Curate submission
Submission: BIGSdb_20200723194830_018868_29257 Summary type: alleles submitte: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-23 status: pending locus: NEIS0001 (IpxC) sequences: 3 technology: Illumina read length: 100-199 coverage: 20-49x assembly: de novo assembly: oftware: Spades
Sequences Messages Identifier Length Sequence Complete CDS Status Query Assigned allele UK322 924 ATGCTGCAAAGAACTTTGGG ACATCAGAAACTTGTGGATAA assigned 1992 UK332 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA assigned 1992 UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAACTTGTGGATAA assigned 1992 Archive Archive Archive Archive of submission and any supporting files:
Action CLOSE SUBMISSION

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.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
	💄 Help 📝 🚯
Curate submission	
Guidie Subillission	
Submission: BIGSdb 20200723194830 018868 29257	
Subilitission. BIGSub_20200/25194050_010000_29257	
Summary	
type: alleles	
submitter: Joe Bloggs, University of Oxford, UK	
datestamp: 2020-07-23	
status: pending	
outcome: accepted - data uploaded	
locus: NEIS0001 (lpxC)	
sequences: 3 ras	
technology: Illumina	
read length: 100-199	
coverage: 20-49x	
assembly: de novo	
assembly software: Spades	
Sequences	Messages
Identifier Length Sequence CDS Status Query Assigned allele	
UK322 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA 🗸 rejected 🗸 🔍 🖋 Curate	
UK323 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA 🗸 rejected 🔍 Q 🖉 Curate	Message: Append Send now
UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA 🗸 rejected 🗸 🖉 Curate	Archive
Mark all: Pending Rejected Update	Archive of submission and any supporting files:
BATCH CURATE BATCH QUERY	Archive of submission and any supporting mes.
	Download TAR
	-

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.

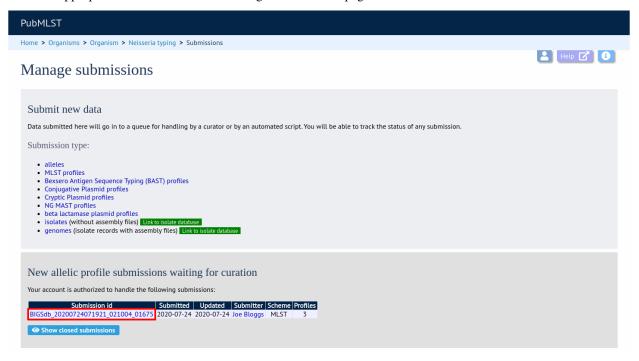
7.1.5 Closing the submission

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	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	E Help 🖌 🕄
Submission: BIGSdb_20200723194830_018868_29257	
type: alleles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-23 status: pending locus: NEIS0001 (lpxC) sequences: 3 technology: Illumina read length: 100-199 coverage: 20-49x assembly: de novo assembly software: Spades	
Sequences Identifier Length Sequence Complete CDS Status Query Assigned allele UK322 924 AT6CTGCAAAGAACTTG6GC ACATCAGAACTTG6GGATAA rejected Q rejected Q Mark all: Pending Rejected Q Mark all: Pending Rejected Q Archive Archive Archive Archive of submission and any supporting files: 2020-07-23 21:42:41+00 Keith Jolley These sequences are not NEIS0001 alleles. Download Tak Message: Append Send now Send now Send now Send now	Action CLOSE SUBMISSION

7.2 Profiles

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



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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	🛓 Help 🔽 🚯
Submission: BIGSdb_20200724071921_021004_01675	
type: profiles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending BATCH CURATE Narkall: Pending Rejected Update	
Messages Archive Archive of submission and any supporting files: Download Download The Message: Append Send now Send now	

7.2.1 Individual profile curation

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	Hetp 🕜 🕄
Submission: BIGSdb_20200724071921_021004_01675	
SummaryProfiles	
type: profiles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending 2020-07-24 Mark all: Pending Rejected Update	
Messages Archive Archive fsubmission and any supporting files:	
Download TAR Message: Append Send now	

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

PubMLST Home > Organisms > Organism > Neisseria typing > Add new profile	
Add new MLST profile	Late C
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record ST: 15581 abc2: 43 adk: 2 aroE: 12 fumC: 32 gdh: 32 pdhC: 3	
pgm: 2 sender: ! Bloggs, Joe (jbloggs) ~ clonal_complex: curator: ! Keith Jolley (keith) date_entered: ! 2020-07-24 datestamp: ! 2020-07-24 PubMed ids: Action	

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.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new profile	
Add new MLST profile	L Help
ST-15581 added. Don't forget to close the submission!	

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.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	Heip 了 🕄
Submission: BIGSdb_20200724071921_021004_01675	
type: profiles UK32 43 submitter: Joe Bloggs, University of Oxford, UK UK33 7 5 datestamp: 2020-07-24 UK34 76 5 status: pending 0	5 3 87 43 34 pending Q Curate Mark all: Pending Rejected Update
Messages Archive Archive Download Archive Arch	NTE

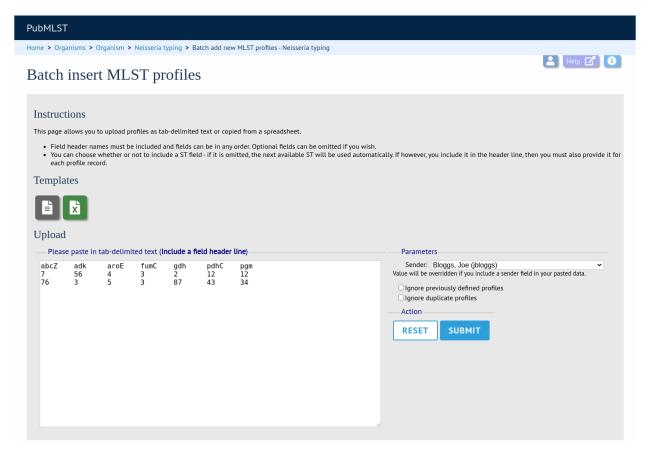
7.2.2 Batch profile curation

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

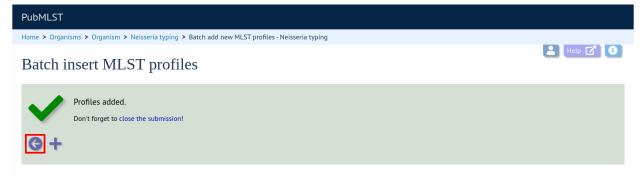
PubMLST		
Home > Organisms > Organism > Neisseria typing > Submissions > C	urate submission	
Curate submission		💄 Help 📝 🚺
Submission: BIGSdb_20200724071921_02100	4_01675	
Summary	Profiles	
type: profiles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending	Identifier abcZ adk aroE fumC gdh pdhC pgm Status Query Assigned ST UK32 43 2 12 52 3 2 assigned Q 15581 UK33 7 56 4 3 2 12 pending ~ Q Curate UK34 76 3 5 87 43 4 pending ~ Q Curate BATCH CURATE Wark all: Pending Rejected Update	
Messages Archive Archive of submiss Download	ision and any supporting files:	
Message: Append Send now		

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

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



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



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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > Cu	Curate submission
Curate submission	
Submission: BIGSdb_20200724071921_021004	04_01675 Profiles
type: profiles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending	Identifier abcZ adk aroE fumC gdh pdhC pgm Status Ouery Assigned ST UK32 43 2 12 32 32 3 2 assigned Q 15581 UK33 7 56 4 3 2 12 12 assigned Q 15581 UK34 76 3 5 3 87 43 4 assigned Q 15583 Messages Archive Archive Archive Archive Archive
	Archive of submission and any supporting files: Download
	Message: Append Send now CLOSE SUBMISSION

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.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > 0	
Curate submission	Help 🔀 🕄
Submission: BIGSdb_20200724071921_02100	04_01675
Summary	Profiles
type: profiles submitter Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending outcome: accepted - data uploaded	Identifier abcZ adk arcE fumC gdh pdhC pgm Status Query Assigned ST UK32 43 2 12 32 32 2 rejected > Q Q Curate UK33 7 56 4 3 2 12 12 rejected > Q Q Curate UK34 76 5 5 8 43 34 rejected > Q Q Curate UK34 76 5 5 8 43 34 rejected > Q Q Curate BATCH CURATE Wark all: Pending Rejected Update
	Messages Archive
	Archive of submission and any supporting files:
	Download Tap
	Message: Append Send now

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

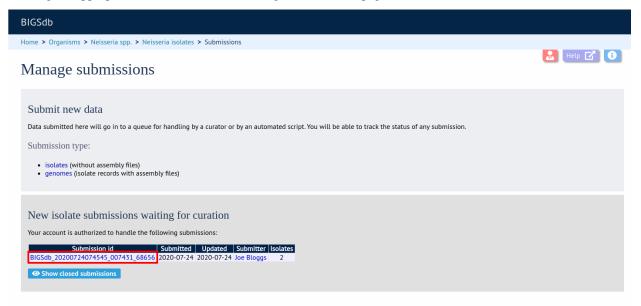
You can add a message to the submitter by entering it in the message box and clicking 'Append'. Once profiles have all been either assigned or rejected, the 'Close submission' button will be displayed. Click this to close the submission.

The submitter will be notified of their submission status.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	Leto 🔽 🚯
Submission: BIGSdb_20200724071921_021004_01675	
Summary Profiles Identifier abcZ adk arcE fum: gdh pdhC pgm Status Query Assigned ST type: profiles UK32 43 2 12 52 3 2 rejected > Q submitter: Joe Bloggs, University of Oxford, UK UK33 7 56 4 3 2 12 rejected > Q UK33 7 56 4 3 2 12 rejected > Q UK34 76 5 5 87 47 3 rejected > Q	
Messages Timestamp User Message 2020-07-24 06:33:44+00 Keith Jolley Please make a corresponding submission to the isolate database. New STs cannot be assigned without this.	
Message: Append Send now	
Archive Action Archive of submission and any supporting files: Download	

7.3 Isolates

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



You will see a table summarizing the submission.

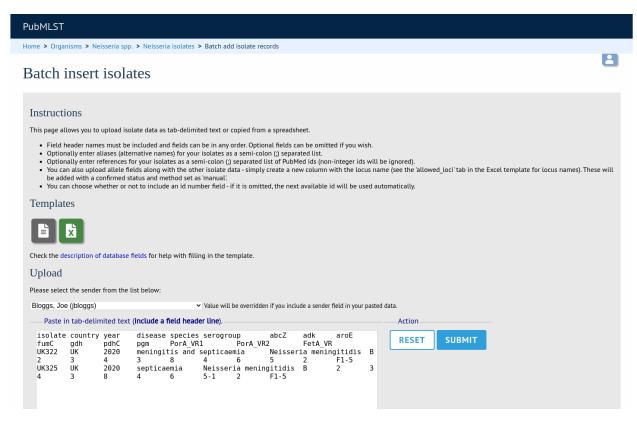
BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Submissions > Curate submission	
Curate submission	Help 🕜 🚯
Submission: BIGSdb_20200724074545_007431_68656 Summary type: isolates submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending	
Isolates Isolate country year disease species serogroup abc2 adk arce fumC gdh pdhC pgm PorA_VR1 PorA_VR2 FetA_VR ST (MLST) UK322 UK 2020 meningitis and septicaemia Neisseria meningitidis B 2 3 4 3 8 4 6 5 2 F1-5 11 UK325 UK 2020 septicaemia Neisseria meningitidis B 2 3 4 3 8 4 6 5-1 2 F1-5 11 Record status: pending v Update Messages Archive Archive of submission and any supporting files:	
Download The Message: Append Send now	

Click the 'Batch curate' button.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria is	olates > Submissions > Curate submission	
Curate submission		
Submission: BIGSdb_202007240 Summary type: isolates submitter: Joe Bloggs, Unive datestamp: 2020-07-24 status: pending		
Isolates Isolate country year disease	species serogroup abcZ adk aroE fumC gdh pdhC pgm PorA_VR1 PorA_VR2 FetA	VR ST (MLST)
UK322 UK 2020 meningitis and septicae UK325 UK 2020 septicaemia		5 11
BATCH CURATE	Record status: pending	
Messages	Archive	
	Archive of submission and any supporting files: Download	
Message: Append Send now		

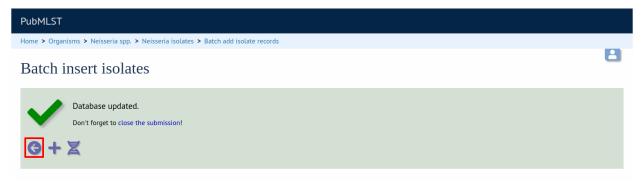
This will take you to the batch isolate upload page in the curators' interface.

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



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

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'.

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

PubMLST	
Home > Organisms > Neisseria spp. > Neisseria isolates > Submissions > Curate submission	
Curate submission	💄 Help 📝 🚯
Submission: BIGSdb_20200724074545_007431_68656	
type: isolates submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending outcome: accepted - data uploaded	
Isolates Isolate country vear disease species iserogroup labcZ adk aroE fumC adh pdhC pam PorA VR1 PorA VR2 FetA VR ST (MLST)	
UK322 UK 2020 meningitis and septicaemia Neisseria meningitidis B 2 3 4 3 8 4 6 5 2 F1-5 11	
UK325 UK 2020 septicaemia Neisseria meningitidis B 2 3 4 3 8 4 6 5-1 2 F1-5 11 Record status: pending V Update	
Messages Archive Action	
Timestamp User Message 2020-07-24 06:59:59+00 Keith Jolley Isolate 'UK322' uploaded - id: 80312. Isolate 'UK325' uploaded - id: 80313. Archive of submission and any supporting files: CLOSE SUBMISSION	
Message: Append Send now	

CHAPTER 8

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
```

(continues on next page)

```
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.
-e, --exemplar
   Only use alleles with the 'exemplar' flag set in BLAST searches to identify
   locus within genome. Specific allele is then identified using a database
   lookup. This may be quicker than using all alleles for the BLAST search,
   but will be at the expense of sensitivity. If no exemplar alleles are set
   for a locus then all alleles will be used. Sets default word size to 15.
-f --fast
    Perform single BLAST query against all selected loci together. This will
   take longer to return any results but the overall scan should finish
   quicker. This method will also use more memory - this can be used with
    --exemplar to mitigate against this.
-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. 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
```

(continues on next page)

(continued from previous page)

```
(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.
--reuse blast
   Reuse the BLAST database for every isolate (when running -- fast option).
   All loci will be scanned rather than just those missing from an isolate.
   Consequently, this may be slower if isolates have already been scanned,
   and for the first isolate scanned by a thread. On larger schemes, such as
   wgMLST, or when isolates have not been previously scanned, setting up the
   BLAST database can take a significant amount of time, so this may be
   quicker. This option is always selected if --new_only is used.
-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 Defining exemplar alleles

Exemplar alleles are a subset of the total number of alleles defined for a locus that encompass the known diversity within a specified identity threshold. They can be used to speed up *autotagging* as the BLAST queries are performed against exemplars to identify the locus region in the genome followed by a direct database lookup of the sequence found to identify the exact allele found. This is usually combined with the autotagger –fast option.

Once exemplars have been defined you may also wish to set the fast_scan="yes" option in the config.xml file. This enables their use for scanning within the web curators' interface.

There is a script called 'find_exemplars.pl' in the BIGSdb scripts/maintenance directory.

A full list of options can be found by typing:

```
find_exemplars.pl --help
NAME
    find_exemplars.pl - Identify and mark exemplar alleles for use
   by tagging functions
SYNOPSIS
   find_exemplars.pl --database NAME [options]
OPTIONS
--database NAME
   Database configuration name.
--datatype DNA|peptide
   Only define exemplars for specified data type (DNA or peptide)
--exclude_loci LIST
    Comma-separated list of loci to exclude
--help
   This help page.
--loci LIST
   Comma-separated list of loci to scan (ignored {f if} -s used).
--locus_regex REGEX
   Regex for locus names.
--schemes LIST
   Comma-separated list of scheme loci to scan.
--update
   Update exemplar flags in database.
--variation DISSIMILARITY
   Value for percentage identity variation that exemplar alleles
    cover (smaller value will result in more exemplars). Default: 10.
```

8.3 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.
--allow_subsequences
   Allow definition of sub- or super-sequences. By default these will not
   be assigned.
-A, --alignment INT
   Percentage alignment (default: 100).
-B, --identity INT
   Percentage identity (default: 99).
-c, --coding_sequences
   Only return complete coding sequences.
```

(continues on next page)

```
-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.
--type_alleles
   Only use alleles with the 'type_allele' flag set to identify locus.
   If a partial match is found then a full database lookup will be performed
```

(continues on next page)

(continued from previous page)

```
to identify any known alleles. Using this option will constrain the search
space so that allele definitions don't become more variable over time. Note
that you must have at least one allele defined as a type allele for a locus
if you use this option otherwise you will not find any matches!
-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.
```

8.4 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

8.5 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 {\tt if} isolate already has sequences {\tt 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.
```

CHAPTER 9

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.

BIGSdb								
Home > Organisms > Organi	sm > Neisseria typing	9						
<i>Neisseria</i> typ	<i>Neisseria</i> typing database							
	Query a sequence					➡) LOG IN		
	Sequence query		Batch sequence q	uery				
	Query a single seque genome assembly to matches.		Query multiple indep in FASTA format to in matches.			Ł DOWNLOADS	-	
Find alleles						Allele sequences Allelic profiles		
	Sequence attribut	search		uence attribute		EXPORT	+	
	Find alleles by matc loci together)	ning criteria (all Select, analys	Select, analyse and c alleles from a single	and download specific single locus.	MALYSIS	+		
		Search for a	llelic profiles				+	
Allelic profile quer		Search by combir	ations of alleles	Batch profile query		() INFORMATION	+	
Search, browse or ent			rtial matches to find	Lookup multiple allelic p	profiles			
		related promes.		togenen.				

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

lGSdb	
ome > Organisms > Organism > Neisseria typing > Download alleles	
Download allele sequences	Help 🔽
Select loci by scheme Alphabetical list All loci by scheme Click within the tree to display details of loci belonging to schemes or groups of schemes - clicking a group folder will display the loci for all schemes within nodes to expand/collapse.	n the group and any subgroups. Click the
III loci III loci IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	

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.

BIGSdb	1						
ome > O	rganism	s > Org	ganism	> Neisseria typin	g > Down	oad alleles	
Dow	nloa	ıd a	llele	e sequen	ces		Help 🔽
	hin the	tree to d	lisplay o	tical list All loci b details of loci belo		hemes or groups o	of schemes - clicking a group folder will display the loci for all schemes within the group and any subgroups. Click the
	Strair Capsu Gene Geno Linea Metal N. gol Plasn Typin Gypin Gypin Gypin	ule tic Inform mic islan ge Sche bolism norrhoea nids	nds :mes ae AMR g antige				
	Downloa						name/product Curator(s) Last updated
abcZ adk	*	DNA	1077 841	Fixed: 433 bp Fixed: 465 bp	433 464	434 465	0. Harrison 2020-06-23 0. Harrison 2020-07-14
aroE	1			Fixed: 490 bp	464	493	0. Harrison 2020-07-14 O. Harrison 2020-06-12
fumC	1		1184	Fixed: 465 bp	463	467	O. Harrison 2020-06-12
gdh	*		1156	Fixed: 501 bp	501	513	0. Harrison 2020-07-03
pdhC	*			Fixed: 480 bp	480	503	0. Harrison 2020-06-04
pgm	*	DNA	1172	Fixed: 450 bp	447	489	O. Harrison 2020-06-12
È	X						

Click the download link for the required locus

MLS	-								
Locus	Download	Туре	Alleles	Length (setting)	Min length	Max length	Full name/product	Curator(s)	Last updated
abcZ	*	DNA	1077	Fixed: 433 bp	433	434		O. Harrison	2020-06-23
adk	*	DNA	841	Fixed: 465 bp	464	465		O. Harrison	2020-07-14
aroE		DNA	1118	Fixed: 490 bp	489	493		O. Harrison	2020-06-12
fumC	*	DNA	1184	Fixed: 465 bp	463	467		O. Harrison	2020-06-12
gdh	4	DNA	1156	Fixed: 501 bp	501	513		O. Harrison	2020-07-03
pdhC	*	DNA	1096	Fixed: 480 bp	480	503		O. Harrison	2020-06-04
pgm	*	DNA	1172	Fixed: 450 bp	447	489		O. Harrison	2020-06-12

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

>fumC_	1
--------	---

(continues on next page)

	(continued from previous page)
ACCATCGGCATGGCGGGCGCGCGGGGCAATTTCGAGCTGAACGTCTATATGCCCGTTATC	
GCCTACAACCTCTTGCAATCCATCCGCCTCTTGGGCGACGCGTGCAACAGCTTCAACGAA	
CACTGCGCCATCGGCATCGAACCCGTACCGGAAAAAATCGACTATTTCCTGCACCATTCC	
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAC	
>fumC_3	
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG	
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGG	
TTGGGCGAAATCAAAATCCCCCGAAAACGAGCCGGGTTCGTCCATCATGCCGGGCAAAGTC	
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT	
ACCATCGGCATGGCGGGCGCGCCGGGGCAATTTCGAGCTGAACGTCTATATGCCCGTTATC	
GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA	
CACTGCGCCGTCGGCATCGAACCCGTACCGGAAAAAATCGACTATTTCCTGCACCATTCC	
CTGATGCTGGTTACTGCGTTAAAACCGTAAAATCGGCTACGAAAAC	

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.

GSdb									
me > Organisms	> Organis	im > N	leisseria t	typing > Download alleles					
ownload	l alle	ele s	sequ	ences					Help 🗹
Select loci by sche	me i Alpha	betical	list All	loci by scheme					
	ine (Aprie	Detieut		toer by serience					
2									
Locus Downle	oad Type	Alleles	Leno	th (setting) Min length Ma	ax length Fu	ull name/produ	ct Last updated		
235 rRNA 📥	DNA	1101			2966				
23S_rRNA 📥	DNA	1101			2966		2019-08-23		
235_rRNA 🛃	DNA	1101			2966				
	DNA		Variable		2966 Min length	Max length		Curator(s)	Last
A Locus	Download	d Type	Variable	: No limits set 2884	Min	Max	2019-08-23	Curator(s)	updated
Locus ArsR [NEIS1769]	Download	d Type DNA DNA	Variable Alleles 195 394	: No limits set 2884 Length (setting)	Min length	Max length	2019-08-23 Full name/product	Curator(s)	updated 2020-05-22 2020-03-20
Locus ArsR [NEIS1769]	Download	d Type DNA DNA	Variable Alleles 195	: No limits set 2884 Length (setting) Variable: No limits set	Min length 270	Max length 291	2019-08-23 Full name/product ArsR family transcriptional regulator	Curator(s)	updated 2020-05-22 2020-03-20
A Locus ArsR [NEIS1769] AsnC [NEIS1566] abcZ		d Type DNA DNA DNA	Variable Alleles 195 394	: No limits set 2884 Length (setting) Variable: No limits set Variable: No limits set	Min length 270 459	Max length 291 471	2019-08-23 Full name/product ArsR family transcriptional regulator transcription regulator AsnC ABC transporter ATP-binding protein		updated 2020-05-22 2020-03-20 2020-06-23
A Locus ArsR (NEIS1769) AsnC [NEIS1566] abcZ abcZ [NEIS1015]	Download	d Type DNA DNA DNA DNA DNA	Alleles 195 394 1077	: No limits set 2884 Length (setting) Variable: No limits set Fixed: 435 bp	Min Length 270 459 433	Max length 291 471 434	2019-08-23 Full name/product ArsR family transcriptional regulator transcription regulator AsnC		updated 2020-05-22 2020-03-20 2020-06-23 2020-07-10
A Locus ArsR (NEIS1769) abc2 abc2 abc2 (NEIS1015) aceF (NEIS1279) ackA2	Download	d Type DNA DNA DNA DNA DNA	Variable Alleles 195 394 1077 1263	Length (setting) Variable: No limits set Variable: No limits set Fixed: 433 bp Variable: No limits set Variable: (1563 min; 1641	Min length 270 459 433 1890	Max length 291 471 434 1944	2019-08-23 Full name/product ArsR family transcriptional regulator transcription regulator AsnC ABC transporter ATP-binding protein		updated 2020-05-22 2020-03-20 2020-06-23 2020-07-10 2020-07-10
A ArsR (NEIS1769) AsnC (NEIS1566) abcZ abcZ (NEIS1015) aceF (NEIS1279) ackA2 [NEIS1727]	Download A A A A A A A A A A A A A	I Type DNA DNA DNA DNA DNA DNA	Variable Alleles 195 394 1077 1263 1550 762	Example 2884 Length (setting) Variable: No limits set Fixed: 453 bp Variable: No limits set Variable: No limits set Variable: (1653 min; 1641 max) Variable: No limits set	Min length 270 459 433 1890 1533 1197	Max length 291 471 434 1944 1641 1212	2019-08-23 Full name/product ArsR family transcriptional regulator transcription regulator AsnC ABC transporter ATP-binding protein dihydrolipoamide acetyltransferase (EC 2.3.1.12) acetate kinase		updated 2020-05-22 2020-03-20 2020-06-23 2020-07-10 2020-07-10 2020-07-10
A Locus ArsR [NEI51769] AsnC [NEI51566] abc2 abc2 [NEI51015] aceF [NEI51279] ackA2 [NEI51727] acnA [NEI51729]	Download A A A A A A A A A A A A A	I Type DNA DNA DNA DNA DNA DNA DNA	Alleles 195 394 1077 1263 1550 762 1211	No limits set 2884 Length (setting) Variable: No limits set Variable: No limits set	Min length 270 459 433 1890 1533 1197 2592	Max length 291 471 434 1944 1641 1212 2616	2019-08-23 Full name/product ArsR family transcriptional regulator transcription regulator AsnC ABC transporter ATP-binding protein dihydrolipoamide acetyltransferase (EC 2.3.1.12) acetate kinase aconitate hydratase 1 (EC 4.2.1.3)		updated 2020-05-22 2020-03-20 2020-06-23 2020-07-10 2020-07-10 2020-06-17 2020-06-23
A Locus ArsR [NEIS1769] abcZ abcZ abcZ [NEIS1015] aceF [NEIS1279] ackA2	Download	t Type DNA DNA DNA DNA DNA DNA DNA DNA	Variable Alleles 195 394 1077 1263 1550 762	Example 2884 Length (setting) Variable: No limits set Fixed: 453 bp Variable: No limits set Variable: No limits set Variable: (1653 min; 1641 max) Variable: No limits set	Min length 270 459 433 1890 1533 1197	Max length 291 471 434 1944 1641 1212	2019-08-23 Full name/product ArsR family transcriptional regulator transcription regulator AsnC ABC transporter ATP-binding protein dihydrolipoamide acetyltransferase (EC 2.3.1.12) acetate kinase		

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.

BIGSdb													
Home > Org	ganism	1s > 0	rganism	> Nei	sseria typing	> Downloa	ad alleles						
Down	ıloa	ad a	allel	e se	equen	ces							Help 🗹
Select loc	i by sc	heme I	Alphabe	tical li	st All loci b	v scheme							
						,							
MLST													
Locus Do	ownloa	id Type	Alleles	Lengt	h (setting)	Min length N	4ax length	Full name/p	oroduct Curator(s) La	ast updated			
abcZ	*		1077		d: 433 bp	433	434		O. Harrison 2				
adk	4		841		d: 465 bp	464	465		O. Harrison 2				
aroE	*		1118		d: 490 bp	489	493		O. Harrison 2				
fumC	4		1184		d: 465 bp	463	467		O. Harrison 2				
gdh	*		1156		d: 501 bp	501	513		O. Harrison 2				
pdhC			1096		d: 480 bp	480	503		O. Harrison 2				
pgm	*	DN/	1172	Fixe	d: 450 bp	447	489		O. Harrison 2	020-06-12			
Finety Locus	ping Dow			Alleles	Length	(setting)	Min length	Max length	I Full name/product	Curator(s)	Last updated		
PorA_VR1			peptide			lo limits set	9	31	PorA variable region		2020-06-30		
PorA_VR2	2		peptide	988		lo limits set	5	41	PorA variable region		2020-06-15		
FetA VR			peptide			lo limits set	20	69	roix variable region		2020-06-08		
rplF sp	pecie	es	Alleles	Lengt		Min Length N 413		Ful	l name/product nal protein L6 fragmer	Curator(s)	Last updated		

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.

abcZ			1077	Fixed: 433 bp	433	434	Full name/product		2020-06-23
adk	*	DNA		Fixed: 465 bp	464	465		O. Harrison	2020-07-14
aroE	*	DNA	1118	Fixed: 490 bp	489	493		O. Harrison	2020-06-12
fumC	*	DNA	1184	Fixed: 465 bp	463	467		O. Harrison	2020-06-12
gdh		DNA	1156	Fixed: 501 bp	501	513		O. Harrison	2020-07-03
pdhC	*	DNA	1096	Fixed: 480 bp	480	503		O. Harrison	2020-06-04
pgm	*	DNA	1172	Fixed: 450 bp	447	489		O. Harrison	2020-06-12

9.2 Scheme profile definitions

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

BIGSdb

Home > Organisms > Organism > Neisseria typing

Neisseria typing database

	Query a sequence								
Sequence query		Batch sequence q	uery						
Query a single sequ genome assembly to matches		Query multiple indep in FASTA format to in matches.			🛃 DOWNLOADS	-			
maches.		materies.			Allele sequences				
	Find	alleles			Allelic profiles				
Sequence attribut		Locus-specific seg search	<u>uence attribute</u>		EXPORT	+			
Find alleles by mate loci together)	ning criteria (all	Select, analyse and o alleles from a single			MALYSIS	+			
	Search for a	llelic profiles	lolic profiles			+			
All Prove fly and			Det la section		+				
Allelic profile query Search, browse or enter list of profiles		nations of alleles artial matches to find	Batch profile que Lookup multiple a together.		SISOLATES				
	retated profiles.		togethel.						

If there is only one scheme available, the link will directly download the profiles. If multiple schemes are available, the link will take you to an intermediate page from where you can select the scheme to download.

BIGSdb Home > Organisms > Organism > Neisseria typing > Download scheme profiles

Download scheme profiles

Name	Download	Profiles	Description	Curator(s)	Last updat
MLST	*	15,577		Odile Harrison, Keith Jolley	2020 07-1
rplF species	*	211		James Bray, Ellie Watkins, Julia Bennett, Kanny Diallo	2019 02-0
Bexsero Antigen Sequence Typing (BAST)	*	3,755	Scheme designed for the surveillance of the antigen targets included in Bexsero® (fHbp, PorA, NHBA and NadA) for protective immunity. It is based on deduced peptide sequence variants.	Charlene Rodrigues	2020 07-1
Conjugative Plasmid	*	242		Ana Cehovin	2018 04-2
Cryptic Plasmid	*	96		Ana Cehovin	201 05-1
GC OMV peptide typing	*	2,283	OMV peptide typing based on most abundant components of gonococcal OMVs	Anastasia Unitt, Odile Harrison	2020 03-3
N. gonorrhoeae cgMLST v1.0	*	8,453			2020 07-1
N. meningitidis cgMLST v1.0	*	18,679	This is a core genome scheme containing 1605 loci found to be present in >=95% N. meningitidis isolates.	Holly Bratcher	2020 07-1
NG MAST	*	19,141	N. gonorrhoeae typing scheme using porB and tbpB fragments. In development, please DO NOT USE		2020 06-2
NG STAR	*	1,933	Neisseria gonorrhoeae Antimicrobial Sequence Type In development, please DO NOT USE	Odile Harrison	202 07-1
OMV peptide typing	*	4,046	OMV peptide typing based on the components of the MeNZB vaccine	Hannah Chan, Charlene Rodrigues	201 05-1

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

ST	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	clonal_complex
⊥ ⇔II	Ţ	3	Ţ	T	Ţ	Ţ	3	ST-1 complex/subgroup I/
2 ⇔II	1	3	4	7	1	1	3	ST-1 complex/subgroup I/
								(continues on next page)

								(continued from previous page)
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_
∽II	I							
6	1	1	2	1	3	2	11	ST-5 complex/subgroup_
∽II	I							
7	1	1	2	1	3	2	19	ST-5 complex/subgroup_
∽II	I							
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
⇔CC	mplex							
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

CHAPTER 10

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.

ull informa	tion on i	solate M10 2	240474 (id:1	18968)		Help
			× ×	,		
🔳 Projects						
This isolate is a member	of the following p	roject:				
The MRE Meningoroccu	s Genome Library	is a collaboration between		coccus Genome Library he Scottish Haemophilus, Legionella, M	eningococcus and Pneumor	
Reference (SHLMPR) La	ooratory, The Well		e and the University of Ox	ford, funded by the Meningitis Research		Meningitis A
Provenance	/primary me	etadata				
	id: 18968		disease:	invasive (unspecified/other)	update history:	223 updates show details
	ate: M10 24047	F 5-11: F3-9: ST-269	species: serogroup:	Neisseria meningitidis B	date entered:	2012-02-15 2020-02-28
-	(cc269)		genogroup:		uatestamp:	2020-02-20
	try: UK [England ent: Europe]	capsule group:	B ERR086224 www.ebi.ac.uk		
reg	on: South East			Dorothea Hill, University of Oxford,		
y epidemiological y	ear: 2010 ear: 07/2010-06	/2011	curator	UK Auto Tagger		
Secondary						
✗ Vaccines						
Bexsero reactiv	ity: cross-reactiv	re 🕕	Trumenba reactivity:	cross-reactive	Trumenba notes:	fHbp_peptide: 15 is cross-reactive to vaccine variant - data derived
Bexsero no	Inotes Inotes Inotes Inotes	e: 15 is cross-reactive		Inotes		from MEASURE assays (PMID:29535195), and SBA assays
bexsero ne	to vaccine v	ariant - data derived				(PMID:22569484, PMID:22718089,
		4709, PMID:23588089,				PMID:23352429, PMID:26407272, PMID:26707218, PMID:26803328,
		998, PMID:26950303, 628, PMID:28366725,				PMID:26835974, PMID:27745812, PMID:27846061, PMID:28196734,
	PMID:30135 PMID:31770	218, PMID:30592763, 063)				PMID:28566335, PMID:29236639)
Publication		,				
Brehony C,	Rodrigues CM, Bor			16). Distribution of Bexsero® Antigen S	equence Types (BASTs) in ir	wasive meningococcal disease
• Hill DM, Lu	idarme J, Gray SJ,		y C, Harrison OB, Bray JE, J	olley KA, Bratcher HB, Parkhill J, Tang C		015). Genomic epidemiology of age-
Rodrigues 0	MC, Chan H, Vipor	d C, Jolley K, Harrison OB,	Wheeler J, Whiting G, Feav	ort study. <i>Lancet Infect Dis</i> 15 :1420-8 vers IM, Maiden MCJ (2018). Typing com		es to understand diversity and
Rodrigues 0	MC, Lucidarme J, I		n JC, Moxon ER, Maiden M	CJ (2018). Genomic Surveillance of 4CM	lenB Vaccine Antigenic Varia	ants among Disease-Causing Neisseri
meningitidi	s Isolates, United	Kingdom, 2010-2016. <i>Eme</i>	rg Infect Dis 24: 675-682	3066 isolates		
Sequence b	n					
	igs: 275	N50 contig number:		ontig number: 97		
	pth: 2,195,045 b pth: 109,859 bp	N50 length (L50): N90 contig number:		length (L95): 3,513 loci tagged: 2,179		
	th: 7,982 bp	N90 length (L90):				
Show sequence	bin					
E Schemes an	d loci					
All loci		Navigate and select sch	emes within tree to displa	ay allele designations		
🛺 Capsule						
🚚 Genetic Inform	-					
Uneage Sche						
🛺 Metabolism						
🛺 Pilin						
🖳 🛄 Other scheme						
Loci not in sc	iemes					
I Tools						

Each record will have some or all of the following sections:

10.1.1 Projects

E Projects	
This isolate is a member of the following project:	
MRF Meningococcus Genome Library	
The MRF Meningococcus Genome Library is a collaboration between Public Health England, the Scottish Haemophilus, Legionella, Meningococcus and Pneu Reference (SHLMPR) Laboratory, The Wellcome Trust Sanger Institute and the University of Oxford, funded by the Meningitis Research Foundation. Use of the MRF Genome Library data must be cited in any publication or presentation making use of it.	umococcus Meningitis Research Foundation

This displays a list of projects that the isolate is a member of. Only projects that have a full description and the 'isolate_display' flag in their settings will be displayed.

10.1.2 Provenance metadata

Provenance/pr	imary metadata				
id:	18968	disease:	invasive (unspecified/other)	update history:	223 updates show details
isolate:	M10 240474	species:	Neisseria meningitidis	date entered:	2012-02-15
strain designation:	B: P1.19-1,15-11: F3-9: ST-269	serogroup:	В	datestamp:	2020-02-28
	(cc269)	genogroup:	В		
country:	UK [England]	capsule group:	В		
continent:	Europe	ENA run accession:	ERR086224 www.ebi.ac.uk		
region:	South East	sender:	Dorothea Hill, University of Oxford,		
year:	2010		UK		
epidemiological year:	07/2010-06/2011	curator:	Auto Tagger		

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.

BIGSdb

Home > Organisms > Neisseria spp. > Neisseria isolates > Isolate information

Full information on isolate M10 240474 (id:18968)

Update history

Timestamp	Curator	Action
2020-02-28 12:18	Auto Tagger	NEIS1428: new designation '63' (sequence bin scan)
		NEIS1428: sequence tagged. Seqbin id: 4957; 28130-30895 (sequence bin scan)
		NEISp1428: new designation '2' (sequence bin scan)
		NEISp1428: sequence tagged. Seqbin id: 4957; 28130-30892 (sequence bin scan)
2020-02-28 12:01	Keith Jolley	NEISp1428: designation '2' deleted
2020-02-28 11:55	Keith Jolley	NEIS1428: designation '63' deleted
2019-12-09 14:41	Odile Harrison	NG_porB: designation '34' deleted
2019-08-18 17:48	Auto Tagger	NEIS1338: new designation '1' (sequence bin scan)
		NEIS1338: sequence tagged. Seqbin id: 4882; 13145-13327 (sequence bin scan)
2019-08-18 17:45	Keith Jolley	NEIS1338: designation '1' deleted
2019-02-16 08:16	Auto Tagger	NG-MAST_porB: new designation '741' (sequence bin scan)
		NG-MAST_porB: sequence tagged. Seqbin id: 4865; 6602-7091 (sequence bin scan)
2018-12-06 12:57	Holly Bratcher	NEIS2072: sequence tagged. Seqbin id: 4917; -282-311 (sequence bin scan)
2018-10-16 09:36	Auto Tagger	NEIS2583: new designation '71' (sequence bin scan)
		NEIS2583: sequence tagged. Seqbin id: 4940; 10070-10305 (sequence bin scan)
		NEIS3171: new designation '12' (sequence bin scan)
		NEIS3171: sequence tagged. Seqbin id: 4965; 28192-28251 (sequence bin scan)
2018-06-26 10:05	Odile Harrison	NEIS2169: designation '6' deleted
2018-06-16 09:30	Auto Tagger	NEIS2669: new designation '63' (sequence bin scan)
		NEIS2669: sequence tagged. Seqbin id: 4818; 22479-22907 (sequence bin scan)
		NEIS2788: new designation '62' (sequence bin scan)
		NEIS2788: sequence tagged. Segbin id: 4913: 18651-19082 (sequence bin scan)

10.1.3 Publications

Publications (4)

- Brehony C, Rodrigues CM, Borrow R, Smith A, Cunney R, Moxon ER, Maiden MC (2016). Distribution of Bexsero® Antigen Sequence Types (BASTs) in invasive meningococcal disease isolates: Implications for immunisation. Vaccine 34:4690-7 2016 Isolates
- Hill DM, Lucidarme J, Gray SJ, Newbold LS, Ure R, Brehony C, Harrison OB, Bray JE, Jolley KA, Bratcher HB, Parkhill J, Tang CM, Borrow R, Maiden MC (2015). Genomic epidemiology of ageassociated meningococcal lineages in national surveillance: an observational cohort study. Lancet Infect Dis 15:1420-8 899 isolates
- Rodrigues CMC, Chan H, Vipond C, Jolley K, Harrison OB, Wheeler J, Whiting G, Feavers IM, Maiden MCJ (2018). Typing complex meningococcal vaccines to understand diversity and population structure of key vaccine antigens. *Wellcome Open Res* 3:151 3505 Isolates
- Rodrigues CMC, Lucidarme J, Borrow R, Smith A, Cameron JC, Moxon ER, Maiden MCJ (2018). Genomic Surveillance of 4CMenB Vaccine Antigenic Variants among Disease-Causing Neisseria meningitidis Isolates, United Kingdom, 2010-2016. Emerg Infect Dis 24:673-682 3066 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



This section contains basic statistics describing the sequence bin. Clicking the 'Show sequence bin' button navigates to the *sequence bin record*.

🔒 Help 📝

10.1.5 Scheme and locus data

A hierarchical tree displays available schemes. Click within internal nodes to expand them.



Clicking any terminal node will display data available for a scheme or group of schemes.

🛺 All loci									
🖙 🛄 Capsule					M	LST			
🛺 Genetic Information Processing	abcZ			fumC		pdhC	pgm		clonal complex
🛺 Genomic islands	4 S	10 S	15 S	9 S	<mark>8</mark> S	11 S	9 S	269	ST-269 complex
🛺 Lineage Schemes									
🛺 Metabolism									
🛺 Pilin									
🚛 🚛 Typing									
🔊 MLST									
···· 🚚 165									
🚚 Antigen genes									

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.

BIGSdb	
Home > Organisms > Organis	m > Neisseria typing > Allele information
Allele informa	tion - abcZ: 2
Provenance/mo locus: allele: sequence:	abcZ 2 TITGATACCG TIGCCGAAGG TITGGGTGAA ATTCGCGGATT TATTGCGCCCG TTACCACCGC GTCGGCATG AGTTGGAAAA CGGTTCGGGT GAGGCTITGT TGAAAGAACT CAACGAATTA CAACTTGAAA TCGAAGCGGAA GGACGGCTGG AAGCTGGATG CGGCAGTCAA GCAGACTTTG GGGGAACTCG GTTGCCCGGA AAACGAAAAA ATCGGCAAGCC TITCCGGCGG TCAGAAAAAG CGTGTGCCT TGGCGCAGGC TTGGGTGCAG AAGCCGACG ATTGCTGCT GGACGAACCG
date entered: datestamp: sender:	Sanger trace checked
Profiles contai MLST:	ning this allele 1262 profiles
🗧 Isolate databas	es
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. 11940 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

BIGSdb	
Home > Organisms	
NEIS0346 allele sequence: id-18968	
Contig position	Options
sequence bin ld: 4956 contig length: 22993 start: 13864 end: 14607 length: 744 orientation: forward complete: yes method; Illumina	Flanking sequence length: 100 ~ Reload
Sequence	
CACTGTACGC GGCGCACGGC TTCAGCATTG CGGGCAGGCG GAAAAACTAT TACCGTACAG CCGACGGTAA AACCGAAGAT GCCGTCTTAA TGGAGAA CCACCTGCAC GAAGCCTTGG GTTTGGGCCC GATGTGGTTG AAGCGGGGGG CCGTCGTCAT GCCGTCCCCA ACATTGCCGG AAAGCCCGAC ACAATC CAGCATTCCG CAGCGTCCGC CCGAACAGCA TACCGGTCAG GCACGGCTCA AAACCATGAA AGCTGTTGGAA ACAACGCCG TACATACGGC CAAACCC GTCCGGCGTT TCAGACGGCA TCGCCCCGT TCCCGCCGCT TCGGGCATAA CCAAACTTGC CGTGTCGC CGACGAGCGC TACATACGGC CAAACCC GCGGCGGTGT CTGCTCGACA ATATACTCAA AGCCGTAGGG CTGCAGGCC CCTATGTCCA CAAAACCTGT TGGGTGAAAA CCGCCGCGCT CGGCAAC GCGAAATGGC CTGGAACAAA CCGCCCGGCA ATCGGACGGC TGCCGCGCGC CCTATGTCCA CAAAACCTGT TGGGTGAAA CCGCCGCGC TGCGCAGCGC GGCCGCCCC TTCTTCATCA TCGACCATCC CGCCCGGCT GTGCGCCGCGCC	ICGC CCGCAAAAGC AAACCGTCCT IGCG CCTGAAACCG AAACGGCTCT TTAC GGGCAACTGT TCCACGGAAA ICCC ATGCCGTCTG AAGCGGCAAT IGCG ATGATTGAAA CTTTGTGCGC IGTC TTGGCGCAAG GCGGCGGCAG
H C T R R T A S A L R A G G K T I T V Q P T V K P K M P S W R K Y F1 T V R G A R L Q H C G Q A E K L L P Y S R R N R R C R L N G E N F2 L Y A A H G F S I A G R R K N Y Y R T A D G K T E D A V L M E K I F3 1 CACTGTACCCGCGCGCGCGCTCAGCATTGCGGCAGGCGGAAAACTATTACCGTACAGCCGAGGTAGAACCCAGGTCTTAATGGAGAAAAT 100 	
K P D T N P P A K A N R P Q H S A A S V R T A Y R S G T A Q N H E F1 S P T Q I R P Q K Q T V L S I P Q R P S E Q H T G Q A R L K T M K F2 K A R H K S A R K S K P S S A F R S V R P N S I P V R H G S K P K F3 201 AAACCCC5ACAAACTAACCCGCAAGCAAACCCATCCGCAAGCATCAGCGCAGCGCTCAGAACCCTAACCCTAAACCCTAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAACCCTAACCATACCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAACCCTAAACCCTAACCCTAACCCTAACCATACCCTAAACCCTAACCCTAACCCTAAACCCTAA	
S V G N N R R T Y A Q T R A M N R N A S V R R F R R H R P R S R R F F1 V L E T T A V H T R K P A P E T E T P L S G V S D G I A P V P A A F2 C W K Q P P Y I R A N P R L K P K R L C P A F Q T A S P P F P P L F3 301 AGTTIGGAAACAACGCGCGAAACCGGCAGCGCGCGCGGAAACGGCATCCGGCGCTTCGGGCGGTTCGGGCGCGCGC	

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



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

BIGSdb	spp. > Neisseria isolates > Isolate inf	o > Sequence bin		
		s sequence on		🔒 Help 🗹 🕄
Sequence bin fo	or M10 240474			
Contig summary stat Contigs: 2 Total length: 2 Minimum length: 1 Maximum length: 2 Maan length: 7 N50 contig number: 2 N50 contig number: 7 N90 contig number: 7 N90 length (L90): 6 N95 contig number: 9 N95 length (L95): 3	130-1 130-1 275 110-1 2,195,045 90-1 145 80-1 109,859 60-1 7,982 40-1 20 30-1 34,308 10-1 100-10-10 500 97 500	ize distribution	Cumulative contig length	
FASTA EMBL GBK	• al Q 4816 4836	4869		x Guide) (Track Labels) (Save SVG) 🗘
promoter region pro_NEIS0488		ри	o_NEIS1635 pro_NEIS0350	¢
partial CDS serC aspA carB		I III I III por8 adk gytA fumC penA pd	fHbp_DNAfrag_Pasteur	↓ ₩4
NEISp0612 NEISp2198	NEISp0715 NEISp1364 NEISp0408	NEISp1130 NEISp0210 NEISp1783 NEISp037 NEISp NEISp1063	NEIS _P 0138 Htbp_peptide	503 NEISp0531 NEISp1687
intergenic region	igr_up_NEIS2109	igr_up_NEIS0055	Http_PEPTDEtrag_Pasteur ig_up_NEIS0351 ig_up_USEIS0350	\$
miscellaneous R_Regulon IRNA-leu IRNA25a IRNA-pro	II I II IRNA-arg IRNA-arg NG_ponA	IRNA-leu IRNA-leu IRNA-asrg	s 'mt/R IIIIIIIIIIIIIIII s 'mt/R IRNA-asp IRNA-typ IRNA-phe IRNA-g III IRNA-thr	II IIII IRNA-gin hmbR II IRNA-gin
Sequence Sequencing method 4869 Illumina	J Original designation NODE_90_length_109787_cov_40.98	Length Comments Locus 3086 109859 NEIS11 NEIS11 NEIS11 NEIS11 NEIS11	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	Annotation EMBL GBK

A sequence bin record contains information about the 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

Charts show the distribution of contig sizes and the cumulative contig length against contig number giving a breakdown indication of contig size.

The record includes an embedded genome browser showing the positions of any loci that have been tagged.

There are also links to download the contigs in FASTA, Genbank or EMBL format, along with annotation in GFF3 format.

Finally there is a table that shows the loci that are tagged on each contig. Individual contigs can also be downloaded in EMBL or Genbank format.

CHAPTER 11

Querying data

11.1 Querying sequences to determine allele identity

Sequence queries are performed in the sequence definition database. Click 'Single sequence' query from the contents page.

BIGSdb Home > Organisms > Organism > Neisseria typing Neisseria typing database

Query a sequence	Find alleles	Search for allelic profiles	D LOG IN	
Single sequence	<u>By specific criteria</u>	By specific criteria		
Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	L DOWNLOADS	
			EXPORT	
Batch sequences Query multiple independent sequences in FASTA format to identify allelic matches.	By locus Select, analyse and download specific alleles from a single locus.	By allelic profile	MALYSIS	
		This can include partial matches to find related profiles.		
		In a batch		
		Look up multiple allelic profiles together.		

Paste your sequence in to the box - there is no need to trim. Often, you can leave the locus setting on 'All loci' - the software should identify the correct locus based on your sequence. Sometimes, especially in databases with a large number of defined loci, 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'.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Sequence query	
Sequence query	Help 🔽 🚺
Please paste in your sequence to query against the database. Query sequences will be checked firs nearest partial matches will be identified if an exact match is not found. You can query using either — 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
ACGGTCAGAT TCCCGATATT CCGCCCGATA TTTTTCAAAA CTGCGCGCTT GCCTACGATA TGGTGTACGG CTGCGCGCA AAACCGTTTT TAGATTTTGC ACGACAATCG GGTGCGGAAAA AAACTGCCGA CGGACTGGGT ATGCTAGTCG GTCAAGCGGC GGCTTCCTAC GCCCTCTGGC GCGGATTTAC GCCCGATATC CGCCCCGTTA TCGAATACAT GAAAGCCCTA TAAACATGTT CCGCATCATC AAATGGCTGA TTGCCCTGCC CGTCGGCATC TTTATCTTTT TCAATGCCTA TGTGTACGGC AACATCATTA CCTACCGG	Select FASTA file: Choose file No file chosen or enter Genbank accession Action RESET SUBMIT

If an exact match is found, this will be indicated along with the start position of the locus within your sequence.

BIGSdb
Home > Organisms > Organism > Neisseria typing > Sequence query
Sequence query
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 nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences.
MLST
Select FASTA file: Choose file No file chosen or enter Genbank accession RESET SUBMIT
1 exact match found. Translate query
Locus Allele Length Contig Start position End position Flags Comments aroE 2 490 Query 97 586

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.

BIGSdb
Home > Organisms > Organism > Neisseria typing > Sequence query
Sequence query
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 nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences.
MLST V locus V Enter query sequence (single or multiple contigs up to whole genome in size) Alternatively upload FASTA file
Select FASTA file: Choose file No file chosen or enter Genbank accession Action RESET SUBMIT
Translate query Closest match: aroE: 64
Show alignment
Differences
2 differences found. 🕥
⁵⁵ A → ¹⁵¹ T ¹³⁴ A → ²³⁰ C
The locus start point is at position 97 of your query sequence. ()

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.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Sequence query	
Sequence query	
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 nearest 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	
All loci v locus v	
Enter query sequence (single or multiple contigs up to whole genome in size) Alternatively upload FASTA file	
Select FASTA file: Choose file No file chosen or enter Genbank accession Action	
RESET SUBMIT	

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.

BIGSdb
Home > Organisms > Organism > Neisseria typing > Sequence query
Sequence query
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 nearest 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 v locus v
Enter query sequence (single or multiple contigs up to whole genome in size) Alternatively upload FASTA file
Select FASTA file: Choose file No file chosen
or enter Genbank accession — — Action
RESET SUBMIT
7 exact matches found.
Locus Allele Length Contig Start position End position Flags Comments
abcZ 2 433 187414 5637 6069
adk 3 465 187432 8935 9399 aroE 4 490 187449 4826 5315
fumC 3 465 187396 4871 5335
gdh 8 501 187466 7333 7833 pdhC 4 480 187542 56831 57310
pgm 6 450 187395 21997 22446
Matching profile
ST: 11
clonal complex: ST-11 complex

11.2 Querying multiple sequences to identify allele identities

You can also query mutiple sequences together. These should be in FASTA format. Click 'Batch sequences' from the contents page.

BIGSdt) ganisms > Organism > Neisseria typing				
Ne	<i>isseria</i> typing databas	se			
	Query a sequence	Find alleles	Search for allelic profiles	DIG IN	
	Single sequence	By specific criteria	By specific criteria		
	Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	Ł DOWNLOADS	+
	matches.			EXPORT	+
	Batch sequences	<u>By locus</u>	By allelic profile	MALYSIS	+
	Query multiple independent sequences in FASTA format to identify allelic matches.	Select, analyse and download specific alleles from a single locus.	This can include partial matches to find related profiles.		+
			In a batch	1 INFORMATION	+
			In a Datch Look up multiple allelic profiles together.	S ISOLATES	
			ugenen		

Paste your sequences (FASTA format) in to the box. Select a specific locus, scheme or 'All loci'.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Batch sequence query	
Batch sequence query	8
Please paste in your sequences 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 nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences.	
Enter query sequences (FASTA format) Alternatively upload FASTA file	
>sample 1 GAAGCGAAAAAAATCATTGACGAAGGCGGCTTGGTGCGCGACCACATCATTATCGGCATGGTCAAAGAAC GCATCGCGCAAGACGACTGCAAAAAACGGTTTCCTGTTGCACGGGTTTCCCGGCGCACATTGGCACAAGCCGA AGCGATGGTTGAAGCAGGCGTGGATTTGGATGCGCTGAAAATCGACGTGCCTGACAACGCGG GACCGTATGAAGCCGCGCCGC	

The best match will be displayed for each sequence in your file. If this isn't an exact match, the differences will be listed.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Batch sequence query	
Batch sequence query	0
Please paste in your sequences 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 nearest 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 v locus v	
Enter query sequences (FASTA format) Alternatively upload FASTA file	
<pre>>sample_1 GAAGCGAAAAAAATCATTGACGAAGGCGGCTTGGTGCGGACGACATCATTATCGGCATGGTCAAAGAAC GCATCGCCGAAGACGACGGCTTGCTGGACGGCGACGACATCATTATCGGCATGGCCAAAGACCGA AGCCGATGGTTGAAGCAGCGCGTGGATTTGGATGCAGCGTTGCCCGCGCACATGGCCACAGCGTGATGGCCGACGACGACGGCGACGACGACGGCGACGACGGCGACGA</pre>	
ContigMatchLocusAlleleDifferencessample_1exactadk3sample_2partialadk211 difference found.237T \rightarrow 237Asample_3partialadk272differences found.230G \rightarrow 230A425C	

11.3 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.

11.3.1 General (all loci) sequence attribute search

To retrieve specific allele designations, click 'By specific criteria' in the 'Find alleles' section on a sequence definition database contents page.

BIGSdt Home > Or) ganisms > Organism > Neisseria typing				
Ne	<i>isseria</i> typing databas	e			
	Query a sequence	Find alleles	Search for allelic profiles	⇒ LOG IN	
	Single sequence	<u>By specific criteria</u>	By specific criteria		
	genome assembly to identify allelic loci together)	Search, browse or enter list of profiles	L DOWNLOADS	+	
	matches.			Z EXPORT	+
	Batch sequences	<u>By locus</u>	By allelic profile		+
	Query multiple independent sequences in FASTA format to identify allelic matches.				+
			la e hatel		+
			In a batch Look up multiple allelic profiles together.		

Enter your query using the dropdown search box - additional terms can be added by clicking the '+' button.

Designations can be queried using standard operators.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Sequence attribute search	
Sequence attribute search	Help 🕜 🕚
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 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 un	
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 settings.	
Search criteria Display	
locus • = • abcZ • i Order by: locus • ascending • Display: 25 • records per page ()	
Filter query by Action	

Click 'Search'.

BIGSdb				
Home > Organisms	> Organism >	Neisseria typing > Sequen	ce attribute search	
Sequence	attribut	te search		Help 🔀 🚺
Some loci have ad export plugins.	lditional fields w	hich are not searchable from	n this general page. S	Search for these at the locus-specific query page. Use this page also for access to the sequence analysis or
		tabase have allele ids defin allele ids using the drop-de		eries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your
Please enter your settings.	search criteria be	elow (or leave blank and su	omit to return all reco	ords). Matching sequences will be returned and you will then be able to update their display and query
Search crite	ria			Display
Combine search				Order by:
locus	✓ =	✓ abcZ	+ (j	locus v ascending v Display: 25 v records per page (j)
allele id	♥ =	v 5		bispus). 25 v lecolus per page ()
	у by	Action		
		RESET	EARCH	
			EARCH	
1 record returned.	Click the hyperli	nk for detailed information		
locus allele id		sequence	sequence length	type allele comments flags
abcZ 5 TTT		C TCGTCGAACTCGA		

Click the hyperlinked results to display allele records.

BIGSdb						
Home > Organisms > Organism > Neisseria typing > Sequence attribute search						
Sequence attribute search						
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						
Combine searches with: AND V Order by:						
locus v = v abcZ + 1 locus v ascending v						
allele id v = v 5 Display: 25 v records per page ()						
- Filter query by Action						
RESET SEARCH						
1 record returned. Click the hyperlink for detailed information.						
Iocus allele id sequence sequence length abcZ S1 TTTGATACCGTTGCC TCGTCGAACTCGATC 433						

BIGSdb		
Home > Organisms > Organis	m > Neisseria typing > Allele information	
Allele informa	tion - abcZ: 5	Help 🔽
length: status: date entered: datestamp:	abcZ 5 TITGATACCG TIGCCGAAGG TITGGGCGAA ATTCGCGATT TATTGCGCCG TTATCATCAT GTCAGCCATG AGTTGGAAAA TGGTTCGAGT GAGGCTTTGT TGAAAGAGCT TAACGAATTG CAACTTGAAA TCGCAAGCGAA GGACGGCTGG AAGCTGGATG GGCGACTCAA GCAGACTTTG GGTGAACTTG GTTGCCAGA AAACGAAAAA ATCGGCAACC TCTCCGCGCG ACGAAAAAAG CGTGTTGCCC TAAGCCAGGC TTGGGTGACATG GGTGAACTTG GTTGCCAGAC ACACAAAAA ATCGGCACTGA CGCGACTGAAAAAC GCGTTGCCT TAGCGCAGGC TTGGGTGAATAC CCACGACCGC CGTTTTTGG ACAACTAATACGCGCCATC GTCGAACTCG ATC 433 Sanger trace checked	
Profiles contai	Man-Suen Chan, University of Oxford ning this allele 302 profiles	
🛢 Isolate databas	ses	
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 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 p sample. 1578 isolates	

Various search criteria can also be selected by combining with filters. Click the filter heading to display these.

BIGSdb
Home > Organisms > Organism > Neisseria typing > Sequence attribute search
Sequence attribute search
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
allele id v < v10 + i Order by: locus v ascending v Display: 25 v records per page (j
Filter query by Action
locus: abcZ status: ~ type allele: ~ sender: ~ curator: ~ allele flag: ~
9 records returned. Click the hyperlinks for detailed information.

locus	allele id	sequence	sequence length	type allele comments fla	igs
abcZ	1	TTTGATACTGTTGCC TTGTCGAACTCGATC	433		
abcZ	2	TTTGATACCGTTGCC TTGTCGAACTCGATC	433		
abcZ	3	TTTGATACCGTTGCC TTGTTGAACTTGACC	433		
abcZ	4	TTTGATACCGTTGCC TTGTCGAACTCGATC	433		
abcZ	5	TTTGATACCGTTGCC TCGTCGAACTCGATC	433		
abcZ	6	TTTGATACCGTTGCC TTGTCGAACTCGATC	433		
abcZ	7	TTTGATACTGTTGCC TTGTCGAACTCGATC	433		
abcZ	8	TTTGATACCGTTGCC TTGTCGAACTTGACC	433		
abcZ	9	TTTGATACCGTTGCC TTGTCGAACTCGATC	433		

11.3.2 Locus-specific sequence attribute search

Some loci have *extended attribute fields*. To query these, click 'By locus' in the 'Find alleles' section on a sequence definition database contents page.

В	BIGSdb
Hom	ne > Organisms > Organism > Neisseria typing

Neisseria typing database

Query a sequence	Find alleles	Search for allelic profiles	DI LOG IN	
Single sequence	By specific criteria	By specific criteria		
Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	L DOWNLOADS	+
moures.			EXPORT	+
Batch sequences Query multiple independent sequences		MALYSIS	+	
in FASTA format to identify allelic matches.	Select, analyse and download specific alleles from a single locus.	This can include partial matches to find related profiles.		+
		In a batch		+
		Look up multiple allelic profiles together.	S ISOLATES	

Pick the required locus from the dropdown box.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Sequence attribute search (locus-	-specific)	
Sequence attribute search - PorA_VR2		Help 🗹 🚺 🗡
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).		
Allele fields	Display	Action
allele id v = v + G	Order by: allele id v ascending v Display: 25 v records per page (RESET

The fields specific for that locus will be added to the dropdown query boxes.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Sequence attribute search (locus-	specific)	
Sequence attribute search - PorA_VR2		Help 🗹 🚺 🗡
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).		
Allele fields	Order by: allele id ascending v Display: 25 v records per page (j	Action RESET SEARCH

The query form can be modified by clicking the 'Modify form options' tab:

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Sequence	attribute search (locus-specific)	
Sequence attribute search - abc	ΈZ	Help 🔀 🚯 🥢
Locus: abcZ	reload when changed nit to return all records).	
Allele fields	Display	Action
allele id 🔹 👻	Image: Provide the second	RESET SEARCH

A list box can be added by clicking the 'Show' button for 'Allele id list box'.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Sequence attribute search (locus-specific)	
Sequence attribute search - abcZ	Help 🗹 💽 🖊
Locus: abcZ Page will reload when changed	×
Further information is available for this locus.	Modify form parameters
Please enter your search criteria below (or leave blank and submit to return all records).	Click to add or remove additional query terms:
Allele fields Display	Allele fields
allele id v = v + () Order by: allele id v ascending v	Allele id list box
Display: 25 v records per page (• • Filters

Close the form modification tab and you can now enter a list of allele ids for retrieval.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Sequence attribute search (locus-specific)	
Sequence attribute search - abcZ	Help 🕜 🕕 🗡
Locus: abcZ 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). 	
Allele fields Display	
allele id v = v + () 2 3 4 5 Order by: allele id v ascending v Display: 25 v records per page () Action RESET SEARCH	

Various analysis and export options will be available for use on the retrieved sequences. These include FASTA output and *Locus Explorer* analysis.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Sequence attribute search (locus-specific)	
Sequence attribute search - abcZ	Help 🕜 🚯 🖌
Locus: abcZ 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). Allele fields Allele id list Display allele id v = v Allele id list Display Display: 25 v records per page () Action RESET SEARCH	
S records returned. Click the hyperlinks for detailed information. Locus allele id sequence sequence length type allele comments flags abcZ 1 TTTGATACCGTTGCC abcZ 2 TTTGATACCGTTGCC abcZ 3	
 Export: FASTA Table Analysis: Locus Explorer 	

11.4 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 'By specific criteria' in the 'Search for allelic profiles' section.

BIGSdb Home > Or) ganisms > Organism > Neisseria typing				
Ne	<i>isseria</i> typing databas	e			
	Query a sequence	Find alleles	Search for allelic profiles	➡) LOG IN	
	Single sequence	By specific criteria	By specific criteria		
	Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	🛃 DOWNLOADS	+
	matches.			EXPORT	+
	Batch sequences	By locus	By allelic profile	MALYSIS	+
	Query multiple independent sequences in FASTA format to identify allelic matches.	Select, analyse and download specific alleles from a single locus.	This can include partial matches to find related profiles.		+
			h - h - h	1 INFORMATION	+
			In a batch Look up multiple allelic profiles together.	SISOLATES	

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 'Search'.

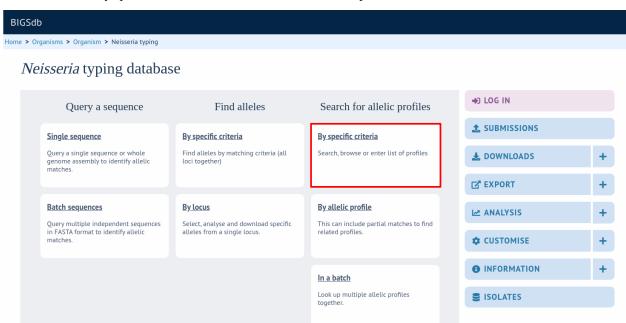
BIGSdb
Home > Organisms > Organism > Neisseria typing > Search or browse profiles
Search or browse profiles
Schemes Please select the scheme you would like to query: MLST
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values. Locus/scheme fields Display/sort options Action ST • = • • • • Order by: ST • ascending • Display: 25 • records per page • • RESET
Browsing all records. 15,577 records returned (1 - 25 displayed). Click the hyperlinks for detailed information. (() () 1 2 3 4 5 6 () () ()
ST abcZ adk aros fumC gdh pdhC pgm clonal complex 1 1 3 1 1 1 3 ST-1 complex 2 1 3 4 7 1 1 5 ST-1 complex 3 1 3 1 1 2 3 ST-1 complex 4 1 3 3 1 1 2 3 ST-1 complex 5 1 1 2 3 ST-4 complex ST-4 complex 5 1 1 2 3 ST-4 complex ST-4 complex 6 1 1 2 3 ST-4 complex ST-5 complex 7 1 1 2 1 3 2 1 ST-5 complex

Clicking the hyperlink for any profile will display full information about the profile.

BIGSdb
Home > Organisms > Organism > Neisseria typing > Profile information
Profile information for ST-11 (MLST)
STabcZadkaroEfumCgdhpdhCpgmclonal complex112343846ST-11 complex
sender: Paula Kriz,
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. 7315 isolates

11.5 Querying scheme profile definitions

Click the link to 'By specific criteria' link in the 'Search for allelic profiles' section.



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'.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Search or browse profiles		
Search or browse profiles		Help 🔀 🚺 🗡
Schemes Please select the scheme you would like to query: MLST v Select		
Enter search criteria or leave blank to browse all records. Modify form parameters to filte Locus/scheme fields Combine searches with: AND × date entered × >= × 2018-03-01 sender (surname) × = × Jolley	er or enter a list of values. Display/sort options Order by: ST v ascending v Display: 25 v records per page ()	Action RESET SEARCH
134 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.		
ST abcZ adk aroE fumC gdh pdhC pgm clonal complex 15607 8 5 55 10 24 17 13608 3 6 9 17 17 21 9 ST-41/44 complex 13609 5 496 17 15 30 650 12 ST-198 complex 13610 7 5 6 13 82 53 15 ST-213 complex 13611 3 6 34 5 36 6 9 ST-41/44 complex 13612 9 6 9 7 9 64 9 ST-41/44 complex 13613 59 39 170 237 148 153 65		

Each field can be queried using standard operators.

Clicking the hyperlink for any profile will display full information about the profile.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Search or browse profiles	
Search or browse profiles	Help 🕜 🚺 🖌
Schemes Please select the scheme you would like to query: MLST Select	
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values. Display/sort options Action Locus/scheme fields Display/sort options Action Combine searches with: AND × Order by: ST × ascending × RESET date entered v >= v 2018-03-01 Display: 25 v records per page () RESET sender (surname) v = v Jolley Order by: ST v records per page () RESET	н
134 records returned (1 - 25 displayed). Click the hyperlinks for detailed information. (*) (*) 1 2 3 4 5 6 (*) (*) ST abcZ adk aroE fumC gdh pdhC ggm clonal complex	
13607 8 5 53 55 10 24 17 13608 3 6 9 17 17 21 9 13609 5 496 17 15 30 650 12 ST-198 complex 13610 7 5 6 13 82 53 15 ST-213 complex 13611 3 634 5 36 6 9 ST-41/44 complex 13612 9 6 9 17 9 64 9 ST-41/44 complex 13613 59 39 170 237 148 153 65	

Other query options are available by clicking the 'Modify form options' tab.

BIGSdb					
Home > Organisms > Org	ganism > Neisseria typing > Search	or browse profiles			
Search or br	rowse profiles				Help 🕜 🚯 🗡
Schemes Please select the schem MLST	ne you would like to query: Select 				
Locus/scheme fiel	leave blank to browse all records. Moc		Display/sort options	Action	
ST	• = •	+ ;	Order by: ST ascending Display: 25 records per page () ascending 	RESET SEARC	н

For example, you can enter a list of attributes to query on by clicking the 'Show' button next to 'Attribute values list'.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Se	arch or browse profiles	
Search or browse profiles		Help 🗹 🚯 🗡
Schemes Please select the scheme you would like to query: MLST Select	I	
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	Modify form parameters
ST v = v	+ () Order by: ST → ascending → Display: 25 → records per page ()	Click to add or remove additional query terms:
		Authorize 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.

BIGSdb						
Home > Organism	s > Organism > Ne	sseria typing > Search or br	owse profiles			
Search o	r browse j	profiles				Help 🕜 🚯 🗡
Schemes Please select the MLST	scheme you would l	ike to query:				
Enter search crite	eria or leave blank to	browse all records. Modify fo	orm parameters to	filter or enter a list of value	5.	
Locus/sche	me fields			Attribute values list	Display/sort options	
ST	¥ =	~	+ ()	Field: ST ~	Order by: ST Display: 25 records pe Action RESET SEARCH	r page
		s for detailed information.				
1 1 3 1	fumC gdh pdhC p 1 1 1	gm clonal complex 3 ST-1 complex				
2 1 3 4		3 ST-1 complex				
3 1 3 1 4 1 3 3		13 ST-1 complex 3 ST-4 complex				
	ools is: BURST rt: Profiles Sequ	ences				

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'.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Search or browse profiles		
Search or browse profiles		Help 🗹 🚯 🗡
Schemes Please select the scheme you would like to query: MLST Select		
Enter search criteria or leave blank to browse all records. Modify form parameters to fi	lter or enter a list of values.	×
Locus/scheme fields	Attribute values list Display/sort	^{optio} Modify form parameters
ST v = v + ()	Field: ST Order by: Display: Action RESET	ST

Available filters will vary depending on the database. These will be combined with other query criteria or lists of attributes.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Search or browse profiles	
Search or browse profiles	Help 🕜 🚯 🗡
Schemes	
Please select the scheme you would like to query:	
MLST V Setect	
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values. Locus/scheme fields ST	
clonal complex: ST-4 complex v () Order by: ST v ascending v Display: 25 v records per page () RESET SEARCH	
1 record returned. Click the hyperlink for detailed information.	
ST abcZ adk aroE fumC gdh pdhC pgm clonal complex 4 1 3 3 1 4 2 3 ST-4 complex	
Analysis tools	
Export: Profiles Sequences	

11.6 Identifying allelic profile definitions

For schemes such as MLST you can query allelic combinations to identify the sequence type (or more generically, the primary key of the profile).

Click the 'By allelic profile' link in the 'Search for allelic profiles' section on a the sequence definition contents page.

BIGSdt Home > Or) ganisms > Organism > Neisseria typing				
Ne	<i>isseria</i> typing databas	e			
	Query a sequence	Find alleles	Search for allelic profiles	→ LOG IN	
	Single sequence	By specific criteria	By specific criteria		
	Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	L DOWNLOADS	+
	matches.			EXPORT	+
	Batch sequences	By locus Select, analyse and download specific	By allelic profile This can include partial matches to find	MALYSIS	+
	Query multiple independent sequences in FASTA format to identify allelic matches.	alleles from a single locus.	related profiles.		+
			In a batch	③ INFORMATION	+
			Look up multiple allelic profiles	SISOLATES	

If multiple schemes are defined in the database you should select the scheme you wish to check.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Search by locus combinations	
Search by locus combinations	Help 🔽 🕄
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	
abcZ adk aroE fumC gdh pdhC pgm ST: Autofill	
Options Display/sort options Action	
Search: Exact or nearest match Order by: ST ascending Display: 25 records per page	

Enter a combination of allelic values (you can enter a partial profile if you wish).

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Search by locus combinations	
Search by locus combinations	Help 🕜 🚯
Schemes Please select the scheme you would like to query: MLST	
Please enter your allelic profile below. Blank loci will be ignored Autofill profile	
abcZ adk aroE fumC gdh pdhC pgm ST: Autofill 2 3 4 3 8 4 6 ST: Autofill	
Options Display/sort options Action	
Search: Exact or nearest match Order by: ST ascending RESET SEARCH SEARCH	

Alternatively, you can automatically populate a profile by entering a value for the scheme primary key field (e.g. ST) and clicking 'Autofill'.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Search by locus combinations	
Search by locus combinations	Help 🕜 3
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	
abcZ adk aroE fumC gdh pdhC pgm ST: 11 Autofil	
Options Display/sort options Action Search: Exact or nearest match v Order by: ST v ascending v Display: 25 v records per page ()	

To find the closest or exact match, leave the search box on 'Exact or nearest match' and click 'Submit'. The best match will be displayed.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Search by locus combinations		
Search by locus combinations		Help 📝 🚯
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	
abcZ adk aroE fumC gdh pdhC pgm 2 3 4 3 8 4 6	ST: 11 Autofill	
Options Display/sort options	Action	
Search: Exact or nearest match Order by: ST ascending ascending Display: 25 records per page	RESET SEARCH	
Exact matches found (7 loci).		
1 record returned. Click the hyperlink for detailed information.		
ST abc2 adk aroE fumC gdh pdhC pgm clonal complex 11 2 3 4 6 ST-11 complex		
Analysis tools		
Export: Profiles Sequences		

Alternatively, if you wish to find all profiles that match the query profile by at least a set number of loci, select the appropriate value in the search dropdown box, e.g. '4 or more matches' will show related profiles that share at least 4 alleles with the query.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Search by locus combinations		
Search by locus combinations		Help 🖌
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	
abcZ adk aroE fumC gdh pdhC pgm 2 3 4 3 8 4 6	ST: 11 Autofill	
Options Display/sort options	Action	
Search: 4 or more matches	RESET SEARCH	
499 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.		
(() () (1 2 3 4 5 6)) (() () () () () () () () () () () ()		
ST abc2/adk atoc fumC gdh pdhC pgm clonal complex 10 2 3 4 2 8 15 2 ST-8 complex 11 2 3 4 3 8 4 6 ST-11 complex 50 2 3 19 8 4 6 ST-11 complex 51 2 3 4 28 6 6 ST-11 complex 52 7 3 4 3 8 4 6 ST-11 complex		

11.7 Batch profile queries

To lookup scheme definitions, e.g. the sequence type for multiple profiles, click 'In a batch' from the 'Search for allelic profiles' section of the sequence definition contents page.

BIGSdb
Home > Organisms > Organism > Neisseria typing

Neisseria typing database

Query a sequence	Find alleles	Search for allelic profiles	DI LOG IN	
Single sequence	By specific criteria	By specific criteria		
Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	Ł DOWNLOADS	-
matches.			EXPORT	-
Batch sequences	By locus	By allelic profile This can include partial matches to find	MALYSIS	
Query multiple independent sequences in FASTA format to identify allelic matches.	Select, analyse and download specific alleles from a single locus.	related profiles.		
		In a batch	() INFORMATION	
		Look up multiple allelic profiles	SISOLATES	

If multiple schemes are defined in the database you should select the scheme you wish to check.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Batch profile query	
Batch profile query	Help 🔽
Schemes Please select the scheme you would like to query: MLST Select	
Enter allelic profiles below in tab-delimited text format using copy and paste (for example directly from a spreadsheet). Columns can be separated by any amount of whitespace. The first column should be an isolate identifier and the remaining columns should comprise the allele numbers (show column order). Click here for example data. Non-numerical characters will be stripped out of the query. Paste in profiles	
Action RESET SUBMIT	

Copy and paste data from a spreadsheet. The first column is the record identifier, and the remaining columns are the alleles for each locus in the standard locus order defined for the scheme. There are links to the column order which can be used as a header line for your spreadsheet and to example data.

Click submit after pasting in the data.

ome > Organisms >	Organism >	Neisseria	typing > E	Batch profile	query						
Batch prof	ile qu	ery									Help 🗹
Schemes Please select the sch	neme you wo	uld like to	query:								
MLST			✓ Select								
column should be ar	isolate iden										
Enter allelic profiles column should be ar stripped out of the q — Paste in profil isolate_1 isolate_2 isolate_3 isolate_4 isolate_5 isolate_5 isolate_6 isolate_7 isolate_8 isolate_10	n isolate iden uery.										
column should be ar stripped out of the q — Paste in profil isolate_1 isolate_2 isolate_3 isolate_4 isolate_5 isolate_6 isolate_6 isolate_7 isolate_8 isolate_9	a isolate iden uery. 2 35 7 6 558 8 222 222 662	6 24 5 43 4 3 1 3	9 15 1 53 39 6 58 58 58 4	5 48 13 3 45 17 275 261 3	9 6 82 74 65 5 30 263 8	6 482 53 8 594 18 3 5 7	8 41 17 54 600 16 255 255 6				

A results table will be displayed.

	nism		rgani	sm >	Neis	seria t	vning	> Bate	h profile query			
ine - orga			gan				, pring	· but	in prome query			Help
Batch	pro	ofil	le (que	ery							
	-			-								
Isolate	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST	clonal complex	Query	2	
isolate 1	2	6	9	5	9	6	8	3021	ST-41/44 complex	Q		
isolate_2	35	24	15	48	6	482	41	7697	ST-192 complex	Q		
isolate_3	7	5	1	13	82	53	17	1997	ST-213 complex	Q		
isolate_4	6	5	53	3	74	8	54	9403		Q		
isolate_5	558	43	39	45	65	594	600	9057	ST-1494 complex (lactamica)	Q		
isolate_6	8	4	6	17	5	18	16	4399	ST-103 complex	Q		
isolate_7	222	3	58	275	30	3	255	8691	ST-4821 complex	Q		
isolate_8	222	1	58	261	263	5	255	11045	ST-4821 complex	Q		
isolate_9	662	3	4	3	8	7	6	11161	ST-11 complex	Q		
isolate_10	12	29	2	26	26	65	17	2040		Q		
isolate_11	9	6	9	9	9	6	653	12841	ST-41/44 complex	Q		
isolate_12	16	2	6	25	17	655	22	10137	ST-53 complex	Q		
isolate_13	4	10	2	5	3	453	9	7008		Q		
isolate_14	3	6	9	60	9	137	9	777	ST-41/44 complex	Q		
isolate_15	28	104	9	9	9	6	9	6747	ST-41/44 complex	Q		

11.8 Investigating allele differences

11.8.1 Sequence similarity

To find sequences most similar to a selected allele within a sequence definition database, expand the 'Analysis' menu item and click 'Sequence similarity' on the contents page.

BIGSdb

Home > Organisms > Organism > Neisseria typing

Neisseria typing database

Query a sequence	Find alleles	Search for allelic profiles	⊕) LOG IN		
Single sequence	By specific criteria	By specific criteria			
Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	🛓 DOWNLOADS	+	
inderes.			Z EXPORT	+	
Batch sequences	By locus	<u>By allelic profile</u>	ANALYSIS	-	
Query multiple independent sequences in FASTA format to identify allelic matches.	Select, analyse and download specific alleles from a single locus.	This can include partial matches to find related profiles.	Sequence similarity Sequence comparison		
		In a batch	Locus Explorer		
		Look up multiple allelic profiles together.		+	
				+	
			SISOLATES		

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.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Plugins > Sequence sin	nilarity	
Find most similar alleles		Help 🔽
This page allows you to find the most similar sequences to a selected allele us	sing BLAST.	
Select parameters	Action	
Locus: abcZ	RESET	

A list of nearest alleles will be displayed, along with the percentage identity and number of gaps between the sequences.

IGSdb													
ome > Organi	isms > Org	ganism > 1	Neisse	ria typing >	Plugins > S	equence	similarity						
ind m	ost si	milaı	r al	leles									Help 🗹
This page allo	ows you to	find the me	ost sim	nilar sequence	es to a selec	ted allele	e using BLA	ST.					
Select p	arameters	s					Act	on					
	Loci	us: abcZ ele: 5				~	RE	SET	SUBMIT				
abcZ-5													
abcZ-5				Alignment (
abcZ-5	99.769	1	0	433/433	বাঁচ								
abcZ-5 Allele % abcZ: 729	99.769 99.769		0		বাঁচ								
abcZ-5 Allete % abcZ: 729 9 abcZ: 671 9	99.769 99.769 99.769	1	0	433/433 433/433 433/433	<u>ब</u> ि बि बि								
abcZ-5	99.769 99.769 99.769 99.769	1 1 1	0 0 0	433/433 433/433 433/433	ৰু কু কু কু কু কু কু কু কু কু কু কু কু কু								
abcZ-5	99.769 99.769 99.769 99.769	1 1 1 1	0 0 0	433/433 433/433 433/433 433/433	<u>क</u> क क क क क								
abcZ-5	99.769 99.769 99.769 99.769 99.769 99.769	1 1 1 1 1	0 0 0 0 0	433/433 433/433 433/433 433/433 433/433	ৰু কু কু কু কু কু কু কু কু কু কু কু কু কু								
abcZ-5 Allete % abcZ: 729 9 abcZ: 671 9 abcZ: 453 9 abcZ: 153 9 abcZ: 166 9 abcZ: 166 9 abcZ: 114 9	99.769 99.769 99.769 99.769 99.769 99.769	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									
abcZ-5 Allete % abcZ: 729 9 abcZ: 671 9 abcZ: 453 9 abcZ: 153 9 abcZ: 166 9 abcZ: 166 9 abcZ: 114 9	99.769 99.769 99.769 99.769 99.769 99.769 99.769 99.769 99.769	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	49 49 49 49 49 49 49 49 49 49 49 49 49 4								

Click the appropriate 'Compare' button to display a list of nucleotide differences and/or a sequence alignment.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Plugins > Sequence comparison	
Allele sequence comparison	Help 🔽
This tool allows you to select two alleles and highlight the nucleotide differences between them.	
Select parameters Action Action	
Locus: abcZ Altele #1: 5 Altele #2: 729 SUBMIT	
Nucleotide differences between abcZ: 5 and abcZ: 729	
Identity: 99.77 %	
Show alignment	
Differences: 1 157: G → T	

11.8.2 Sequence comparison

To directly compare two sequences, expand the 'Analysis' section and click 'Sequence comparison' on the contents page of a sequence definition database.

BIGSdb

Home > Organisms > Organism > Neisseria typing

Neisseria typing database

Query a sequence	Find alleles	Search for allelic profiles	DI LOG IN	
Single sequence	By specific criteria	By specific criteria		
Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	L DOWNLOADS	+
matches.			EXPORT	+
Batch sequences	<u>By locus</u>	By allelic profile	MALYSIS	-
Query multiple independent sequences in FASTA format to identify allelic matches.	Select, analyse and download specific alleles from a single locus.	This can include partial matches to find related profiles.	Sequence similarity Sequence comparison Locus Explorer	
		<u>In a batch</u>	Locus Explorer	
		Look up multiple allelic profiles together.		+
			() INFORMATION	+
			SISOLATES	

Enter the locus and two allele identifiers to compare. Press submit.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Plugins > Sec	juence comparison	
Allele sequence comparison		Help 📝
This tool allows you to select two alleles and highlight the nucleotic	e differences between them.	
Select parameters	Action	
Locus: abcZ v Allele #1: 5 Allele #2: 8	SUBMIT	

A list of nucleotide differences and/or an alignment will be displayed.

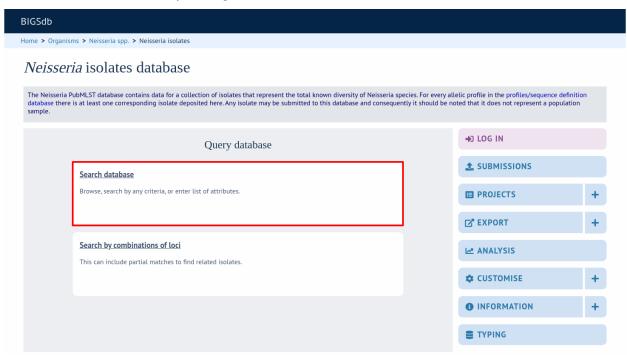
BIGSdb	
Home > Organisms > Organism > Neisseria typing > Plugins > Sequence comparison	
Allele sequence comparison	Help 🗹
This tool allows you to select two alleles and highlight the nucleotide differences between them. Select parameters Action Locus: abcZ V Allele #1: 5 Allele #2: 8	
Nucleotide differences between abcZ: 5 and abcZ: 8 Identity: 90.53 % Show alignment Differences: 41 72: $G \rightarrow T$ 78: $A \rightarrow G$ 79: $A \rightarrow C$ 81: $T \rightarrow C$ 82: $G \rightarrow A$ 83: $G \rightarrow A$ 83: $G \rightarrow A$ 83: $G \rightarrow A$ 87: $G \rightarrow A$ 89: $G \rightarrow A$ 89: $G \rightarrow A$ 99: $G \rightarrow C$	

See also:

Locus explorer plugin.

11.9 Browsing isolate data

Isolate records can be browsed by clicking the link to 'Search database'.



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 'Search'.

	Sdb												
ome	> Organis	ms > Neisseria spp. > Nei	sseria isolates >	Searc	h or brows	e database							
ea	arch c	or browse da	tabase									He	elp 🔽 🚯 🖌
Ente	er search cri	iteria or leave blank to brov	wse all records. N	1odify 1	form param	neters to filter or e	enter a list of values.						
	Isolate nr	ovenance fields					Display/sort options						
							1.						
1	id	✓ =	✓ Enter valu			+ (i	Order by: id		~		✓ as	cending	*
							Display: 25 🗸 re	cords per page	(i)				
	Action												
Г													
	RESET	SEARCH											
66,6	02 records	returned (1 - 25 displayed).	. Click the hyperl	inks fo	r detailed i	nformation.							
66,6	02 records			inks fo	r detailed i	nformation.							
56,6 «	02 records	returned (1 - 25 displayed). 1 2 3 4 5 6		inks fo	r detailed i	nformation.							
66,6 «	02 records			inks fo	r detailed i	nformation.							
66,6	02 records		@ @)	r detailed i fields (i)	nformation.				MLST	Finet	yping anti	gens
«	isolate	1 2 3 4 5 6 aliases	Country	solate 1 year	fields (j	disease	species	capsule group	5 ST		PorA_VR1	PorA_VR2	FetA_VR
«	isolate	1 2 3 4 5 6 aliases B1; NIBSC_2803; Z1001	Country USA	olate 1 year 1937	fields (j) invasive (U	disease Inspecified/other)	Neisseria meningitidis	А	o ST	clonal complex	PorA_VR1 5-2	PorA_VR2 10	FetA_VR F1-5
~A	isolate 4/M1027 120M	1 2 3 4 5 6 aliases	Country USA Pakistan	olate 1 year 1937 1967	fields) invasive (U meningitis	disease Inspecified/other) s and septicaemia	Neisseria meningitidis Neisseria meningitidis	A	1		PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR
« A M	isolate (4/M1027 120M 00242905	1 2 3 4 5 6 aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Country USA Pakistan UK	solate 1 year 1937 1967 2000	fields () invasive (u meningitis invasive (u	disease Inspecified/other) s and septicaemia Inspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B	1 1099	clonal complex ST-1 complex	PorA_VR1 5-2	PorA_VR2 10	FetA_VR F1-5
« A M	Isolate 4/M1027 120M 00242905 M1027	1 2 3 4 5 6 aliases B1; NIBSC_2803; Z1001	Country USA Pakistan UK USA	solate 1 year 1937 1967 2000 1937	fields () invasive (u meningitis invasive (u invasive (u	disease Inspecified/other) s and septicaemia Inspecified/other) Inspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A	1 1099 4	clonal complex ST-1 complex ST-4 complex	PorA_VR1 5-2 5-2 19	PorA_VR2 10 10 15	FetA_VR F1-5
	isolate 4/M1027 120M 00242905 M1027 00240227	1 2 3 4 5 6 aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Country USA Pakistan UK USA UK	Solate 1 year 1937 1967 2000 1937 2000	fields () invasive (u meningitis invasive (u invasive (u invasive (u	disease Inspecified/other) s and septicaemia Inspecified/other) Inspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B	1 1099 4 1100	clonal complex ST-1 complex ST-4 complex ST-32 complex	PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR F1-5
	isolate (4/M1027 120M 00242905 M1027 00240227 00282207	1 2 3 4 5 6 aliases B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043	Country USA Pakistan UK USA UK UK	Solate 1 year 1937 1967 2000 1937 2000 2000	fields) invasive (u meningitis invasive (u invasive (u invasive (u invasive (u	disease inspecified/other) s and septicaemia inspecified/other) inspecified/other) inspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B W	1 1099 4 1100 1101	clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22 complex	PorA_VR1 5-2 5-2 19 7	PorA_VR2 10 10 15 16	F1-5 F5-1
	isolate (4/M1027 120M 00242905 M1027 00240227 00282207 7891	1 2 3 4 5 6 aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Country USA Pakistan UK USA UK UK Finland	Solate 1 year 1937 1967 2000 1937 2000 2000 1975	fields () invasive (u meningitis invasive (u invasive (u invasive (u invasive (u	disease inspecified/other) s and septicaemia inspecified/other) inspecified/other) inspecified/other) inspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A	1 1099 4 1100 1101 5	clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22 complex ST-5 complex	PorA_VR1 5-2 5-2 19	PorA_VR2 10 10 15 16 9	FetA_VR F1-5
	isolate (4/M1027 120M 00242905 M1027 00240227 00282207 7891 00242007	1 2 3 4 5 6 aliases B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043	Country USA Pakistan UK USA UK UK Finland UK	Solate 1 year 1937 1967 2000 1937 2000 2000 1975 2000	fields () invasive (u meningitis invasive (u invasive (u invasive (u invasive (u invasive (u	disease inspecified/other) s and septicaemia inspecified/other) inspecified/other) inspecified/other) inspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A B	1 1099 4 1100 1101 5 1102	clonal complex ST-1 complex ST-32 complex ST-32 complex ST-22 complex ST-18 complex	PorA_VR1 5-2 5-2 19 7	PorA_VR2 10 10 15 16	F1-5 F5-1
	isolate (4/M1027 120M 00242905 M1027 00240227 00282207 7891 00242007 00242007	1 2 3 4 5 6 aliases B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043	Country USA Pakistan UK USA UK UK Finland UK	olate 1 year 1937 1967 2000 1937 2000 1975 2000 1975 2000 1984	fields () invasive (u meningitis invasive (u invasive (u invasive (u invasive (u invasive (u invasive (u	disease inspecified/other) and septicaemia inspecified/other) inspecified/other) inspecified/other) inspecified/other) inspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A B W	1 1099 4 1100 1101 5 1102	clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22 complex ST-5 complex	PorA_VR1 5-2 5-2 19 7	PorA_VR2 10 10 15 16 9	F1-5 F5-1

Clicking the hyperlink for any record will display *full information* about the profile.

lGSdb												
ome > Organis	sms > Neisseria spp. > Nei	isseria isolates >	Searc	h or brow	se database							
Search o	or browse da	tabase									He	łp 🕜 🕒 .
Enter search cr	riteria or leave blank to brow	wse all records. №	lodify f	form parai	meters to filter or e	enter a list of values.						
Isolate n	rovenance fields					Display/sort options						
id	✓ =	✓ Enter valu	10			Order by: id				N 20	cending	
iu		- Enter valu	ie		+ ()	Display: 25 V reco	ords per page			• 43	centuing	
RESET	SEARCH											
	s returned (1 - 25 displayed) 1 2 3 4 5 6		inks for	r detailed	information.							
	s returned (1 - 25 displayed)	@ @)		information.				MLST	Finet	voing anti	gens
66,602 records	s returned (1 - 25 displayed) 1 2 3 4 5 6 aliases	Country	olate f year	ields (j	disease		capsule grou	p ST	MLST clonal complex			FetA_VR
66,602 records	s returned (1 - 25 displayed) 1 2 3 4 5 6 aliases B1; NIBSC_2803; 21001	ls Country USA	olate f year 1937	ields 👔	disease (unspecified/other)	Neisseria meningitidis	А		clonal complex	PorA_VR1 5-2	PorA_VR2 10	F1-5
66,602 records	s returned (1 - 25 displayed) 1 2 3 4 5 6 aliases	Is country USA Pakistan	olate f year 1937 1967	ields 👔 invasive (meningiti	disease (unspecified/other) is and septicaemia	Neisseria meningitidis Neisseria meningitidis	A A	1		PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR
66,602 records	s returned (1 - 25 displayed) 1 2 3 4 5 6 aliases B1; NIBSC_2803; 21001	Is country USA Pakistan UK	olate f year 1937 1967 2000	ields (j) invasive (meningiti invasive (disease (unspecified/other) is and septicaemia (unspecified/other)	Neisseria meningitidis	А		clonal complex	PorA_VR1 5-2	PorA_VR2 10	F1-5
66,602 records (() Isolate 1 A4/M1027 2 120M 3 M00242905 4 M1027 5 M00240227	alisscs B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035	USA USA USA USA UK	olate f year 1937 1967 2000 1937 2000	ields (j) invasive (meningiti invasive (invasive (invasive (disease (unspecified/other) is and septicaemia (unspecified/other) (unspecified/other) (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B	1 1099 4 1100	clonal complex ST-1 complex ST-4 complex ST-32 complex	PorA_VR1 5-2 5-2	PorA_VR2 10 10	F1-5
66,602 records (() () () () () () () () () () () () ()	alisses B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043	Is country USA Pakistan UK USA UK UK	olate f year 1937 1967 2000 1937 2000 2000	invasive (meningiti invasive (invasive (invasive (invasive (disease (unspecified/other) is and septicaemia (unspecified/other) (unspecified/other) (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B W	1 1099 4 1100 1101	Clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22 complex	PorA_VR1 5-2 5-2 19 7	PorA_VR2 10 10 15 16	FetA_VR F1-5 F5-1
66,602 records (() Isolate 1 A4/M1027 2 120M 3 M00242905 4 M1027 5 M00240227 6 M00282207 7 7891	alisscs B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035	Is country USA Pakistan UK USA UK UK Finland	olate f year 1937 1967 2000 1937 2000 2000 1975	ields) invasive (meningiti invasive (invasive (invasive (invasive (disease (unspecified/other) is and septicaemia (unspecified/other) (unspecified/other) (unspecified/other) (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A B A B W A	1 1099 4 1100 1101 5	Clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22 complex ST-5 complex	PorA_VR1 5-2 5-2 19	PorA_VR2 10 10 15 16 9	F1-5
66,602 records (() () () () () () () () () () () () ()	alisses B1; NIBSC_2803; 21001 B35; NIBSC_2822; 21035 B43; NIBSC_3076; 21043	Country USA Pakistan UK USA UK UK Finland UK	olate f year 1937 1967 2000 1937 2000 2000 1975 2000	invasive (meningiti invasive (invasive (invasive (invasive (invasive (invasive (disease (unspecified/other) is and septicaemia (unspecified/other) (unspecified/other) (unspecified/other) (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B A B W	1 1099 4 1100 1101 5 1102	Clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22 complex	PorA_VR1 5-2 5-2 19 7	PorA_VR2 10 10 15 16	FetA_VR F1-5 F5-1

11.10 Querying isolate data

The 'Search database' page of an isolate database allows you to also search by combinations of provenance criteria, scheme and locus data, and more.

BIGSdb								
Home > Organisms >	Neisseria spp. > Neisseria isolates							
Neisseria isolates database								
	.ST database contains data for a collection of isolates that represent the total known diversity of Neisseria sp least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequer							
	Query database		DIG IN					
Se	earch database							
Bro	owse, search by any criteria, or enter list of attributes.		PROJECTS	+				
L			EXPORT	+				
	a <mark>rch by combinations of loci</mark> is can include partial matches to find related isolates.		MALYSIS					
				+				
				+				

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'.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	Help 🗹 🚺 🧨
Enter search criteria or leave blank to browse all records. Modify form parameters to filter	or enter a list of values.
Isolate provenance fields	Display/sort options
Combine with: AND •	Order by: id 🗸 ascending 🗸
country v = v USA + i	Display: 25 🗸 records per page 👔
year • >= • 2000	Action
	RESET SEARCH

After the search has been submitted, the results will be displayed in a table.

BIGSdb											
lome > Organisms > N	eisseria spp. > Ne	isseria isolates > Search or b	rowse database								
Search or b	owse da	tabase								Help	30/
		wse all records. Modify form p									
 Isolate provenar 	ce fields		[isplay/sort op	tions						
Combine with: AND	~			Order by: id					v 8	ascending 🗸	
country	✓ =	✓ USA	+ ()	Display: 25	v re	cords per page ()			5	
year	✓ >=	✓ 2000				in her					
,		2000	/	ction							
				RESET	SEAR	сн					
(<) (<) 1	2 3 4 5 6	Isolate fields (j)				MLST		yping antig			
id isolate alia	ises country year	r disease	species	capsule group	ST	clonal complex	PorA_VR1	PorA_VR2	FetA_VR		
341 M7085	USA 2000) invasive (unspecified/other)				ST-11 complex					
341 M7085 499 MDL01A0601	USA 2000 USA 2001	1 meningitis	Neisseria meningitidis	Y	1378	ST-23 complex					
M7085 499 MDL01A0601 500 MDL01A2447	USA 2000 USA 2001 USA 2001	1 meningitis 1 invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	Y Y	1378 1379	ST-23 complex ST-23 complex					
341 M7085 499 MDL01A0601 500 MDL01A2447 866 MD01227	USA 2000 USA 2001 USA 2001 USA 2001	1 meningitis 1 invasive (unspecified/other) 1	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Y Y	1378 1379 1624	ST-23 complex ST-23 complex ST-167 complex					
341 M7085 499 MDL01A0601 500 MDL01A2447 866 MD01227 867 MD01056	USA 2000 USA 2001 USA 2001 USA 2001 USA 2001 USA 2001	1 meningitis 1 invasive (unspecified/other) 1	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Y Y	1378 1379 1624 1625	ST-23 complex ST-23 complex ST-167 complex ST-23 complex					
341 M7085 499 MDL01A0601 500 MDL01A2447 866 MD01227 867 MD01056 868 MD01066	USA 2000 USA 2001 USA 2001 USA 2001 USA 2001 USA 2001 USA 2001	1 meningitis 1 invasive (unspecified/other) 1 1 1	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Y Y	1378 1379 1624 1625 1626	ST-23 complex ST-23 complex ST-167 complex ST-23 complex ST-269 complex		2	54.4		
341 M7085 499 MDL01A0601 500 MDL01A2447 866 MD01277 867 MD01056 868 MD01066 2281 M7089	USA 2000 USA 2001 USA 2001 USA 2001 USA 2001 USA 2001 USA 2001 USA 2001	meningitis invasive (unspecified/other) I invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Y Y W	1378 1379 1624 1625 1626 11	ST-23 complex ST-23 complex ST-167 complex ST-23 complex ST-269 complex ST-11 complex	5	2	F1-1		
341 M7085 499 MDL01A0601 500 MDL01A2447 866 MD01227 867 MD01056 868 MD01066	USA 2000 USA 2001 USA 2001 USA 2001 USA 2001 USA 2001 USA 2000 USA 2000	1 meningitis 1 invasive (unspecified/other) 1 1	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Y Y W	1378 1379 1624 1625 1626 11 11	ST-23 complex ST-23 complex ST-167 complex ST-23 complex ST-269 complex	5	2 2 3	F1-1		

Each field can be queried using standard operators.

More search features are available by clicking the 'Modify form options' tab on the top of the screen.

BIGS	Sdb													
Home	> Organisms > Neisser	ia spp.	> Neis	seria isolates > Search or b	rowse database									
Sea	arch or brov	vse	dat	abase								He	elp 🔽	8 🗡
Ente	er search criteria or leave	blank	to brow	se all records. Modify form p	arameters to filter or en	iter a list of va	alues.							
	Isolate provenance fie	lds			D) isplay/sort o	options							
Co	ombine with: AND V					Order by:	· · · ·				× 5	ascending	~	
		- =		✓ USA	+ (i)			cords per page G	<u> </u>		• 6	ascending	•	
		· >=		✓ 2000				cords per page (9					
	,			2000	—A	ction								
2,68	32 records returned (1 - 2:			ick the hyperlinks for detaile		RESET	SEAR	СН						
	-													
				Isolate fields 👔				MLST		ping antig				
id	isolate aliases			disease	species	capsule grou		clonal complex	PorA_VR1	PorA_VR2	FetA_VR			
341	M7085 MDL01A0601	USA USA	2000	invasive (unspecified/other) meningitis	Neisseria meningitidis Neisseria meningitidis	W		ST-11 complex ST-23 complex						
	MDL01A0601 MDL01A2447	USA		invasive (unspecified/other)		Y Y		ST-23 complex						
866	MD01227	USA	2001		Neisseria meningitidis			ST-167 complex						
867	MD01056	USA	2001		Neisseria meningitidis			ST-23 complex						
868	MD01066	USA	2001		Neisseria meningitidis			ST-269 complex						
2281	M7089	USA		invasive (unspecified/other)		W		ST-11 complex	5	2	F1-1			
2299		USA		invasive (unspecified/other) invasive (unspecified/other)		W		ST-11 complex ST-22 complex	5	2				
2316	M7086	USA					22		6					

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

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
	Help 🕜 🚯 🖌

Search or browse database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.										×			
	Isolate provenan	ce fields			C	isplay/sort op	tions			Modi	fy form	n parameters	
Co	mbine with: AND	~				Order by: id						- F	
(country	v =		✓ USA	+ ()	Display: 25	5 ¥ re	ecords per page (Click to	add or rem	ove additional query terms:	_
1	/ear	✓ >=		✓ 2000					, I		Provenance	e fields	
					A	ction					Secondary		
											,		
						RESET	SEAR	КСН			Allele desi	gnations/scheme field value	s
										0	Allele desi	gnation counts	
										0	Allele desi	gnation status	
2 6 8	2 records returned	(1 - 25 displ	aved) C	lick the hyperlinks for details	d information						Tanned ser	uence counts	
2,00	2 records returned	(1 - 25 uispu	ayeu). c	2,682 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.									
G			5 6								Tagged sec	quence status	
\ll) (1 2	2 3 4	56	$\odot \odot$							Tagged seo Attribute v	•	
«) (1 2	234	56	$\odot \otimes$						0	55	•	
«) (1 2	2 3 4	56	Isolate fields (j)				MLST	Fine	0	Attribute v	•	
(«) < 1 2 isolate alia	2 3 4 ases country		Isolate fields () disease	species	capsule group	ST	MLST clonal complex		0	Attribute v	•	
341	M7085	ases country USA	year 2000				11	clonal complex ST-11 complex		0	Attribute v	•	
341 499	M7085 MDL01A0601	ases country USA USA	year 2000 2001	disease invasive (unspecified/other) meningitis	Neisseria meningitidis Neisseria meningitidis	W Y	11 1378	clonal complex ST-11 complex ST-23 complex		0	Attribute v	•	
341 499 500	M7085 MDL01A0601 MDL01A2447	ases country USA USA USA	year 2000 2001 2001	disease invasive (unspecified/other) meningitis invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W Y Y	11 1378 1379	clonal complex ST-11 complex ST-23 complex ST-23 complex		0	Attribute v	•	
341 499 500 866	M7085 MDL01A0601 MDL01A2447 MD01227	uses country USA USA USA USA	year 2000 2001 2001 2001	disease invasive (unspecified/other) meningitis invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W Y Y	11 1378 1379 1624	clonal complex ST-11 complex ST-23 complex ST-23 complex ST-167 complex		0	Attribute v	•	
341 499 500 866 867	M7085 MDL01A0601 MDL01A2447 MD01227 MD01056	ases country USA USA USA USA USA	year 2000 2001 2001 2001 2001	disease invasive (unspecified/other) meningitis invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W Y Y	11 1378 1379 1624 1625	Clonal complex ST-11 complex ST-23 complex ST-23 complex ST-167 complex ST-23 complex		0	Attribute v	•	
341 499 500 866 867 868	M7085 MDL01A0601 MDL01A2447 MD01227 MD01056 MD01066	ases country USA USA USA USA USA USA	year 2000 2001 2001 2001 2001 2001	disease invasive (unspecified/other) meningitis invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W Y Y	11 1378 1379 1624 1625 1626	clonal complex ST-11 complex ST-23 complex ST-23 complex ST-167 complex ST-23 complex ST-269 complex	PorA_VR1	0	Attribute v Filters	•	
341 499 500 866 867 868 281	M7085 MDL01A0601 MDL01A2447 MD01227 MD01056 MD01066 M7089	ases country USA USA USA USA USA USA USA	<pre>/ year 2000 2001 2001 2001 2001 2001 2001</pre>	disease invasive (unspecified/other) meningitis invasive (unspecified/other) invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W Y Y	11 1378 1379 1624 1625 1626 11	Clonal complex ST-11 complex ST-23 complex ST-23 complex ST-167 complex ST-269 complex ST-269 complex ST-11 complex	PorA_VR1	0	Attribute v	•	
341 499 500 866 867 868	M7085 MDL01A0601 MDL01A2447 MD01227 MD01056 MD01066	ases country USA USA USA USA USA USA	<pre>/ year 2000 2001 2001 2001 2001 2001 2000 2000</pre>	disease invasive (unspecified/other) meningitis invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W Y Y W W	11 1378 1379 1624 1625 1626	clonal complex ST-11 complex ST-23 complex ST-23 complex ST-167 complex ST-23 complex ST-269 complex	PorA_VR1	0	Attribute v Filters	•	

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.

11.10.1 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.

GSdb		
me > Organisms > Neisseria spp. > Neisseria isolates > Search or br earch or browse database	rowse database	Help 🗹 🔒 🗸
Enter search criteria or leave blank to browse all records. Modify form p Isolate provenance fields id v = Filter value Display/sort options Order by: id Display: 25 v records per page ()	Allele designations/scheme fields	 Ente Modify form parameters Click to add or remove additional query terms: Provenance fields Secondary metadata Altele designations/scheme field values Altele designation status Tagged sequence counts Attribute values list Filters

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 entered in other sections.

BIGSdb									
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database									
Search or brows	e database					Help 🕜 🚺 🗡			
Enter search criteria or leave blar	k to browse all records. Modify form	parameters to filter or ent	ter a list of value	s.					
Isolate provenance fields		AI	llele designatio	ns/scheme fields					
id 🕶 =	✓ Enter value	🛛 🕂 🚯 ST	(MLST)	=	✓ 11	+ ()			
— Display/sort options			Action						
Order by: id		✓ ascending ✓	RESET	SEARCH					
Display: 25 🗸 records	per page 👔	L		SEARCH					
	played). Click the hyperlinks for deta	iled information.							
6,991 records returned (1 - 25 dis,	played). Click the hyperlinks for deta			MLST					
6,991 records returned (1 - 25 dis	played). Click the hyperlinks for deta	species c	capsule group ST						
6,991 records returned (1 - 25 dis, (*) (*) 1 2 3 4 1 I 2 3 4 1 I Isolate allases country 79 160 USA 100 638 USA	blayed). Click the hyperlinks for deta 5 6 () () lsolate fields () year disease 1993 invasive (unspecified/other 1994 invasive (unspecified/other	species c) Neisseria meningitidis) Neisseria meningitidis	capsule group ST C 11 C 11	MLST clonal complex ST-11 complex ST-11 complex					
6,991 records returned (1 - 25 dis) (() () () 1 2 3 4 () 1 2 3 4 () 1 0 638 () 1 0 638 () 1 0 638 () 1 0 54 () 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Isolate fields () year disease 1993 invasive (unspecified/other 1994 invasive (unspecified/other 2000 invasive (unspecified/other	species c) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis	capsule group ST C 11 C 11 C 11	MLST Clonal complex ST-11 complex ST-11 complex ST-11 complex					
6,991 records returned (1 - 25 dis (*) (*) 1 2 3 4 10 isolate allases country 79 160 USA 100 638 USA 156 00-1008 UK [Scotland 156 100-1050 UK [Scotland	Isolate fields () year disease 1993 Invasive (unspecified/other 1904 Invasive (unspecified/other 1900 Invasive (unspecified/other 2000 Invasive (unspecified/other 2000 Invasive (unspecified/other	<mark>species c</mark>) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis	capsule group ST C 11 C 11 C 11 C 11	MLST clonal complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex					
6,991 records returned (1 - 25 dis (*) (*) (*) 1 2 3 4 1 3 1 2 1 3 1 2	Isolate fields () year disease 1993 invasive (unspecified/other 1904 invasive (unspecified/other 2000 invasive (unspecified/other 2000 invasive (unspecified/other 2000 invasive (unspecified/other 2000 invasive (unspecified/other	species c Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group ST C 11 C 11 C 11 C 11 C 11	MLST clonal complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex					
6,991 records returned (1 - 25 dis (() () () () () () () () () () () () ()	Isolate fields () Isolate fields () Isolate fields () Vear disease 1993 invasive (unspecified/other 1994 invasive (unspecified/other 1900 invasive (unspecified/other 1 2000 invasive (unspecified/other 1 2000 invasive (unspecified/other 1 2000 invasive (unspecified/other 2 2000 invasive (unspecified/other 2 2000 invasive (unspecified/other	species c) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis	capsule group ST C 11 C 11 C 11 C 11 C 11 C 11 C 11	MLST Clonal complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex					
6,991 records returned (1 - 25 dis (*) (*) 1 2 3 4 10 isolate allases country 79 160 USA 100 638 USA 156 00-108 UK [Scotland 161 00-122 UK [Scotland 164 00-125 UK [Scotland	Isolate fields () year disease 1993 Invasive (unspecified/other 1994 Invasive (unspecified/other 1900 Invasive (unspecified/other 2000 Invasive (unspecified/other	species c Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group ST C 11 C 11 C 11 C 11 C 11 C 11 C 11 C 1	MLST [clonal complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex					
6,991 records returned (1 - 25 disp (*) (*) 1 2 3 4 1 3 0 0 0 0 0 0 0 1 3 0 0 0 0 0 0 1 4 0 0 0 0 0 1 5 0 0 0 1 5 0 0 0 0 1 5 0 0 1	Isolate fields () Isolate fields () Isolate fields () Vear disease 1993 invasive (unspecified/other 1994 invasive (unspecified/other 1900 invasive (unspecified/other 1 2000 invasive (unspecified/other 1 2000 invasive (unspecified/other 1 2000 invasive (unspecified/other 2 2000 invasive (unspecified/other 2 2000 invasive (unspecified/other	species c) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis	capsule group ST C 11 C 11 C 11 C 11 C 11 C 11 C 11 B 11	MLST Clonal complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex ST-11 complex					

11.10.2 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.

BIGSdb		
lome > Organisms > Neisseria spp. > Neisseria isolates > Se	earch or browse database	
Search or browse database		Help 🔀 🚺 🖋
Enter search criteria or leave blank to browse all records. Mod	lify form parameters to filter or enter a list of values.	×
Isolate provenance fields	Display/sort options	Modify form parameters
id • Enter value.	• • • • Order by: id Display: 25 • records per page •	Click to add or remove additional query terms:
Action		Provenance fields
		Secondary metadata
RESET SEARCH		Allele designations/scheme field values
		Allele designation counts
		Allele designation status
		Tagged sequence counts
		• Tagged sequence status
		Attribute values list
		• Filters

For example, to find all isolates that have designations at >1000 loci, select 'total designations > 1000', then click 'Submit'.

	Sdb												
lome	> Organis	ms > Neisseria spp. > Neisse	eria isolates >	Search or	browse database								
Sea	arch c	or browse data	abase								l	Help 🔽	8
Ent	er search cri	iteria or leave blank to brows	e all records. M	odify form	parameters to filter or e	nter a list of values.							
	Isolate nr	ovenance fields				Allele designation cou	ints						
	id		✓ Enter valu			ount of total designation		> ~ 100	- 00				
	iu	• =	• Enter valu	e	+ () Co	ount of total designation	13	2 · 100		í			
	Display/s	ort options				Action							
	Order b	y: id			✓ ascending ✓								
						RESET SE	ARCH						
	Displa	ay: 25 🖌 records per page (i)										
	Displa	ay: 25 🗸 records per page (Ð										
	Displa	ay: 25 V records per page (D										
27,5		iy: 25 ✓ records per page (returned (1 - 25 displayed). Cl	9	nks for det	tailed information.								
27,5		returned (1 - 25 displayed). Cl	9	nks for det	tailed information.								
27,5			9	nks for det	tailed information.								
27,5		returned (1 - 25 displayed). Cl	9	nks for det	tailed information.								
27,5		returned (1 - 25 displayed). Cl	ick the hyperli)					MLST	Finet	voing anti	gens	
27,5 («		returned (1 - 25 displayed). Cl	ick the hyperli	nks for det) olate field year		species			MLST onal complex		yping anti PorA VR2		
(«	28 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6	ick the hyperli	olate field year	ls (j)	species		up ST clu	MLST onal complex				
(«	528 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases	ick the hyperli	olate field year 1937 inv	ls () disease	species Neisseria meningitidis	capsule grou			PorA_VR1	PorA_VR2	FetA_VR	
(« id	528 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases B1; NIBSC_2803; Z1001	ick the hyperli	olate field year 1937 inv 1967 me	ls () disease rasive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis	capsule grou A A	1 S	onal complex	PorA_VR1 5-2	PorA_VR2 10	FetA_VR F1-5	
id 1 2 7	528 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases B1; NIBSC 2803; Z1001 B35; NIBSC 2822; Z1035	ick the hyperli	olate field year 1937 inv 1967 me 1975 inv	Is ① disease rasive (unspecified/other) ningitis and septicaemia	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule grou A A	1 S 5 S	onal complex T-1 complex	PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR F1-5 F5-1	
id 1 2 7 10	528 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2760; Z1054 B54; NIBSC_2760; Z1054	ick the hyperli () () () () () () () () () ()	olate field year 1937 inv 1967 me 1975 inv 1971 inv	Is () disease asive (unspecified/other) ningitis and septicaemia asive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule grou A A A	1 S 5 S 1 S	onal complex <mark>T-1 complex</mark> T-5 complex	PorA_VR1 5-2 5-2 20	PorA_VR2 10 10 9	FetA_VR F1-5 F5-1 F3-1	
id 1 2 7 10 11	isolate A4/M1027 120M 7891 6748 129E 139M	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases B1; NIBSC_2803; Z1001 B35; NIBSC_2760; Z1054 B54; NIBSC_2760; Z1054 B73; NIBSC_2764; Z1073 B92; NIBSC_2782; Z1092	ick the hyperli country USA Pakistan Finland Canada Germany	olate field year 1937 inv 1967 me 1975 inv 1971 inv 1964 inv 1968	Is () disease asive (unspecified/other) ningitis and septicaemia asive (unspecified/other) asive (unspecified/other) asive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule grou A A A A	1 S 5 S 1 S 1 S	onal complex T-1 complex T-5 complex T-1 complex	PorA_VR1 5-2 5-2 20 18-1	PorA_VR2 10 10 9 3	FetA_VR F1-5 F5-1 F3-1 F5-1	
id 1 2	528 records 528 r	allasos allasos B1; NIBSC 2803; 21001 B35; NIBSC 2803; 21001 B35; NIBSC 2822; 21035 B34; NIBSC 2822; 21035 B34; NIBSC 2822; 21035 B35; NIBSC 2828; 21092 B92; NIBSC 2813; 21213	ick the hyperli country USA Pakistan Finland Canada Germany	olate field year 1937 inv 1967 me 1975 inv 1971 inv 1964 inv 1968 1973 inv	disease asive (unspecified/other) ningitis and septicaemia asive (unspecified/other) asive (unspecified/other) asive (unspecified/other) asive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule grou A A A A A A	1 S 5 S 1 S 1 S 1 S 1 S 4 S	T-1 complex T-5 complex T-5 complex T-1 complex T-1 complex T-1 complex T-4 complex	PorA_VR1 5-2 20 18-1 5-2 5-2 5-2 7	PorA_VR2 10 10 9 3 10	FetA_VR F1-5 F5-1 F3-1 F3-1 F3-6 F5-1 F1-5	
id 1 2 7 10 11 13	isolate A4/M1027 120M 7891 6748 129E 139M	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases B1; NIBSC_2803; Z1001 B35; NIBSC_2760; Z1054 B54; NIBSC_2760; Z1054 B73; NIBSC_2764; Z1073 B92; NIBSC_2782; Z1092	ick the hyperli () () () () () () () () () ()	olate field year 1937 inv 1967 me 1975 inv 1971 inv 1964 inv 1968 1973 inv	Is () disease asive (unspecified/other) ningitis and septicaemia asive (unspecified/other) asive (unspecified/other) asive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule grou A A A A A A A	1 S 5 S 1 S 1 S 1 S 4 S 5 S	T-1 complex T-5 complex T-1 complex T-1 complex T-1 complex T-1 complex	PorA_VR1 5-2 5-2 20 18-1 5-2 5-2	PorA_VR2 10 10 9 3 10 10	FetA_VR F1-5 F5-1 F3-1 F3-1 F3-6 F5-1	

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.

BIGSdb								
Home > Organisms > Neisseria spp. > Neisseria	isolates > Search or	browse database						
Search or browse datab	ase						Help 🗹	()
Enter search criteria or leave blank to browse all	records. Modify form	parameters to filter o	or enter a list of values.					
Isolate provenance fields			Allele designation	counts				
id 🗸 = 🗸	Enter value	+ ()	Count of abcZ		>¥ 1	+ ()		
Display/sort options			Action					
Order by: id Display: 25 V records per page (j)		✓ ascending ✓	RESET	SEARCH				
10 records returned. Click the hyperlinks for deta								
id isolate aliases cou	Isolate fields Intry year	(i) disease	species	capsule group	ST	MLST	Finetyping antig	

									PILSI	Filletyping antigens		
id	isolate	aliases	country	year	disease	species	capsule group	ST	clonal complex	PorA_VR1	PorA_VR2	FetA_VR
451	14/1455	NIBSC_2732; Z4717	Russia	1970	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1
887	OX9930357		UK [England]	1999	carrier	Neisseria meningitidis	В	43	ST-41/44 complex	19	15-1	F1-98
3412	OX9931676		UK [England]	1999	carrier	Neisseria meningitidis	В	1228	ST-41/44 complex	7-1	1	
3415	OX9931776		UK [England]	1999	carrier	Neisseria meningitidis	В	1644	ST-213 complex	22	14	F5-5
4193	OX9931563		UK [England]	1999	carrier	Neisseria meningitidis	C	136	ST-41/44 complex	5	2	F1-20
34733	LNP15075		Burkina Faso	1997		Neisseria meningitidis				20	10-1	F3-1
57890	15134		Spain	2002	invasive (unspecified/other)	Neisseria meningitidis	W			5-1	10-4	F1-18,F1-7
61162	NmissMen5		Italy	2016		Neisseria meningitidis				12	16-11	F1-5
89074	61/2013		Austria	2013	meningitis	Neisseria meningitidis	В	1193,15334	ST-116 complex	21	10-1,16	F1-5,F4-2
91694	NML2009-301		Canada	2009		Neisseria meningitidis	В	154,9411	ST-41/44 complex	7-2	4	F1-18

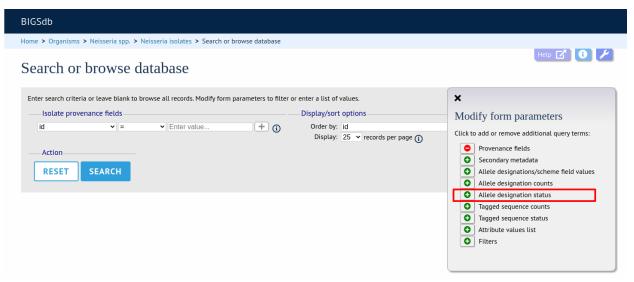
Additional search terms can be combined using the '+' button. Designation count queries will be combined with terms entered in other sections.

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.

11.10.3 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.



Select a locus from the dropdown box and either 'provisional' or 'confirmed'. Additional query fields can be displayed by clicking the '+' button. Click 'Submit'.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	Help 🕜 🚺 🖌
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.	
Isolate provenance fields Allele designation status	
id v = v Enter value + () adk is provisional v + ()	
Display/sort options Action	
Order by: id RESET SEARCH	
Display: 25 v records per page ()	
2 records returned. Click the hyperlinks for detailed information. Isolate fields () MLST Finetyping antigens id isolate aliases country year generation 14/1735 N9703 BennetTfres2 Unknown Neisseria subflava 10304 19363 961-5945 Unknown Neisseria subflava 10304	
Analysis tools	
Reakdown: Fields Two Field Combinations Polymorphic sites Publications Sequence bin	
Analysis: BURST Codons Gene Presence Genome Comparator BLAST rMLST species id PCR	
Export: Dataset Contigs Sequences	
Third party: GrapeTree ITOL PhyloViz Microreact	

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.

11.10.4 Query by annotation status

Isolates can be queried by the annotation status of particular schemes if these have been set up. The idea is that for a well-annotated record the isolate would be expected to have allele designations for all loci in the scheme. Alternatively, different thresholds for number of loci with allele designations can be set up by the scheme administrator to indicate good or bad quality thresholds.

Make sure that the annotation status fieldset is displayed by selecting it in the 'Modify form options' tab.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	Help 🗹 🚯 🗡
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values. Isolate provenance fields Id	Provenance fields

Additional search terms can be combined using the '+' button. Annotation status queries will be combined with terms entered in other sections.

lGSdb											
ne > Organ	iisms > Neisseria spp. > <i>Neisse</i>	eria isolates >	Search or browse	database							
earch	or browse data	abase								Help	2 8 /
Enter search	n criteria or leave blank to brows	e all records. N	lodify form param	neters to filter or er	nter a list of values.						
Isolate	provenance fields				Annotation sta	tus					
id	* =	• Enter	value	+ ()		D • ST • go	od	• +	i		
Display	y/sort options						Acti	on			
Orde	er by: id				- asce	nding -					
Dis	splay: 25 records per page	í			_		RE	SET	EARCH		
51,286 recor	rds returned (1 - 25 displayed). (1 2 3 4 5 6	lick the hyper)	information.						whine anti	
		le le	olate fields 🕥					MICT	Einat		
isolate	aliases		olate fields ()	disease	species	capsule group		MLST	Finet		
		country USA	year year	disease unspecified/other)	species Neisseria meningitidis	capsule group A					
A4/M102 120M	27 B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	country USA Pakistan	year 1937 invasive (u 1967 meningitis	unspecified/other) s and septicaemia	Neisseria meningitidis Neisseria meningitidis		4 S	onal complex T-4 complex T-1 complex	PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR F1-5 F5-1
A4/M102 120M 7891	27 B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B54; NIBSC_2760; Z1054	country USA Pakistan Finland	year 1937 invasive (u 1967 meningitis 1975 invasive (u	unspecified/other) and septicaemia unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A	4 S 1 S 5 S	onal complex T-4 complex <mark>T-1 complex</mark> T-5 complex	PorA_VR1 5-2 5-2 20	PorA_VR2 10 10 9	FetA_VR F1-5 F5-1 F3-1
A4/M102 120M 7891 6748	 B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B54; NIBSC_2760; Z1054 B73; NIBSC_2784; Z1073 	country USA Pakistan Finland Canada	year 1937 invasive (u 1967 meningitis 1975 invasive (u 1971 invasive (u	unspecified/other) s and septicaemia unspecified/other) unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A	4 S 1 S 5 S 1 S	onal complex T-4 complex T-1 complex T-5 complex T-1 complex	<pre>PorA_VR1 5-2 5-2 20 18-1</pre>	PorA_VR2 10 10 9 3	FetA_VR F1-5 F5-1 F3-1 F5-1
A4/M102 120M 7891 6748 1 129E	 B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B54; NIBSC_2760; Z1054 B73; NIBSC_2784; Z1073 B92; NIBSC_2828; Z1092 	country USA Pakistan Finland Canada Germany	year 1937 invasive (u 1967 meningitis 1975 invasive (u 1971 invasive (u 1964 invasive (u	unspecified/other) and septicaemia unspecified/other) unspecified/other) unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A	4 S 1 S 5 S 1 S 1 S	T-4 complex T-4 complex T-1 complex T-5 complex T-1 complex T-1 complex	<pre>PorA_VR1 5-2 5-2 20 18-1 5-2</pre>	PorA_VR2 10 10 9 3 10	FetA_VR F1-5 F5-1 F3-1 F5-1 F3-6
A4/M102 120M 7 7891 0 6748	 B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B54; NIBSC_2760; Z1054 B73; NIBSC_2784; Z1073 	country USA Pakistan Finland Canada Germany	year 1937 invasive (t 1967 meningitis 1975 invasive (t 1971 invasive (t 1964 invasive (t 1968	unspecified/other) s and septicaemia unspecified/other) unspecified/other) unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A	4 S 1 S 5 S 1 S 1 S 1 S	onal complex T-4 complex T-1 complex T-5 complex T-1 complex	<pre>PorA_VR1 5-2 5-2 20 18-1</pre>	PorA_VR2 10 10 9 3	FetA_VR F1-5 F5-1 F3-1 F5-1

11.10.5 Query by sequence bin size and number of contigs

Isolates can be queried based on the total length of sequences within the sequence bin and/or the number of contigs. Make sure that the sequence bin fieldset is displayed by selecting it in the 'Modify form options' tab.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	Help 🔽 🚺 🗡
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values. Isolate provenance fields id	 Modify form parameters Click to add or remove additional query terms: Provenance fields Secondary metadata Allele designations/scheme field values Allele designation counts Allele designation status Sequence bin Tagged sequence counts Attribute values list Filters

Additional search terms can be combined using the '+' button. Sequence bin queries will be combined with terms entered in other sections.

ome >	Organisms >	Neisseria spp. > Neisseria isolates > Search or browse	database									
ane -	organisms >	respected spp respected isotates - bedicit of blowse	. uuuuuase							Help	616	
Sea	rch or l	browse database								Thetp		20
- Cu	i en or .	oromoe autuouse										
Enter	search criteria	or leave blank to browse all records. Modify form parameters	eters to filter or enter	a list of	values.							
b	solate proven	ance fields	[Sequ	ence bin							
id	l	 Enter value 	+ (i)		e with: AND							
						1.8 0 +	()					
			L	numb	er of contigs 📩 <=							
C	Display/sort o					Action						
	Order by: i				 ascending 	RESET	SEARCH					
	Display: 2	25 💽 records per page 🛈										
		_ records per page O										
573 re	ecords returned	d (1 - 25 displayed). Click the hyperlinks for detailed infor	rmation.									
573 re	~	1(1 - 25 displayed). Click the hyperlinks for detailed infor 2 3 4 5 6 () ()	rmation.									
573 re	~		rmation.									
573 re	~	2 3 4 5 6 ()(>)	rmation. Diate fields (j)						MLST	Finet	typing anti	igens
573 re	~	2 3 4 5 6 ()(>)		year	disease	species	capsule group	ST	MLST clonal complex		PorA_VR2	1
«	<i>1</i>	2 3 4 5 6 🔊 🔊	blate fields (j)	year 1999	disease carrier	species Neisseria meningitidis		ST 53			<u> </u>	2 FetA
id	() 1 isolate	2 3 4 5 6 🕥 👀	olate fields (j)		carrier invasive	Neisseria meningitidis Neisseria	group		clonal complex	PorA_VR1	PorA_VR2	2 FetA F5-
id 30 237	 1 isolate 14 	2 3 4 5 6 () () allases BennettTree10; alpha14	olate fields ① country Germany	1999	carrier	Neisseria meningitidis Neisseria meningitidis Neisseria	group cnl	53	clonal complex ST-53 complex	PorA_VR1 7	PorA_VR2 30-3	2 FetA_ F5- F3-
id 30 2377 240	(*) 1 isolate 1 14 1 H44/76 1	2 3 4 5 6 (>) (>) allases BennettTree10; alpha14 44/76-3; NIBSC_2724; Z3842	olate fields ① country Germany Norway	1999 1976	carrier invasive	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria	group cnl B	53 32	clonal complex ST-53 complex ST-32 complex	PorA_VR1 7 7	PorA_VR2 30-3 16	2 FetA F5- F3- F1-
«	I isolate 14 H44/76 MC58	2 3 4 5 6 (>) (>) allases BennettTree10; alpha14 44/76-3; NIBSC_2724; Z3842 BennettTree07; Z7176	olate fields ① country Germany Norway UK	1999 1976 1983	carrier invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	group cnl B B	53 32 74	clonal complex ST-53 complex ST-32 complex ST-32 complex	PorA_VR1 7 7 7 7	PorA_VR2 30-3 16 16-2	1

11.10.6 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.

GSdb					
ne > Organisms > Ne	isseria spp. > 1	Neisseria isolates > Search or	prowse database		Help 🗹 🚺
earch or br	owse d	atabase			
nter search criteria or l	leave blank to b	rowse all records. Modify form	parameters to filter or	enter a list of values.	×
Isolate provenanc	ce fields			Display/sort options	Modify form parameters
id Action RESET SE	▼ =	✓ Enter value	+ 3	Order by: id Display: 25 💙 records per page	 Provenance fields Secondary metadata Allele designations/scheme field values Allele designation counts Allele designation status
					 Tagged sequence counts Tagged sequence status Attribute values list Filters

For example, to find all isolates that have sequence tags at >1000 loci, select 'total tags > 1000', then click 'Submit'.

ome >	Organisa												
	Organish	ns > Neisseria spp. > Neisse	eria isolates >	Search	or browse database								
ear	rch o	or browse data	abase									Help 🔽	
Enter s	search crit	teria or leave blank to brows	e all records. №	1odify fo	orm parameters to filter or e	enter a list of values.							
Ise	solate pro	ovenance fields				Tagged sequence cour	nts						
id			✓ Enter valu	ie		ount of total tags		> • 1	.000 +	(i)			
Di)isplay/sc	ort options				Action				Ŭ			
						riceron							
	1.1	· ·			and the second second								
	Order b	,	~		✓ ascending ✓	RESET SEA	ARCH						
	Order b	y: id y: 25 v records per page (1		✓ ascending ✓	RESET SEA	ARCH						
	Order b Displa	,	iick the hyperli)	detailed information.	RESET	ARCH						
27,521	Order b Displa 1 records r	25 ▼ records per page (eturned (1 - 25 displayed). Cl 1 2 3 4 5 6	ick the hyperli) solate fie	detailed information.				MLST		yping antig		
27,521 ((()))	Order b Displa 1 records r	 25 ▼ records per page (returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases 	iick the hyperli	solate fie	detailed information. elds () disease	species	capsule grou	ир ST	MLST clonal complex	PorA_VR1	PorA_VR2	FetA_VR	
27,521 (() () d is 1 A4/	Order b Displar 1 records r (solate	 25 ▼ records per page (returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allascs B1; NIBSC_2803; Z1001 	ick the hyperli	solate fie year 1937 i	detailed information. elds () disease invasive (unspecified/other)	species Neisseria meningitidis	capsule grou A		clonal complex	PorA_VR1 5-2	PorA_VR2 10	FetA_VR F1-5	
27,521 (() () d is 1 A4/ 2 1	Order b Displar 1 records r (solate (M1027 120M	25 ▼ records per page (eturned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases B1; NIBSC _2803; Z1001 B35; NIBSC _2802; Z105	ick the hyperli	solate fie year 1937 i 1967	detailed information. elds () disease invasive (unspecified/other) meningitis and septicaemia	species Neisseria meningitidis Neisseria meningitidis	capsule grou A A	1	clonal complex ST-1 complex	PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR F1-5 F5-1	
27,521 (() d is 1 A4/ 2 1 7 7	Order b Displa 1 records r (solate (M1027 120M 7891	22 ▼ records per page (returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B54; NIBSC_2760; Z1054	ick the hyperli () () () () () () () () () ()	solate fie year 1937 i 1967 1975 i	elds () disease meningitis and septicaemia musaive (unspecified/other) meningitis and septicaemia	species) Neisseria meningitidis a Neisseria meningitidis) Neisseria meningitidis	capsule grou A A A	1	ST-1 complex ST-5 complex	PorA_VR1 5-2 5-2 20	PorA_VR2 10 10 9	FetA_VR F1-5 F5-1 F3-1	
27,521 (() d is 1 A4/ 2 1 7 7 0 6	Order b Displa 1 records r () () () () () () () () () () () () ()	25 ▼ records per page (eturned (1 - 25 displayed). Cl 1 2 3 4 5 6 aliascs B1; NIBSC 2803; Z1001 B35; NIBSC 2822; Z1035 B54; NIBSC 2786; Z1054 B73; NIBSC 2766; Z1054	ick the hyperli Country USA Pakistan Finland Canada	solate fie year 1937 i 1967 1975 i 1971 i	detailed information. disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule grou A A A A	1 5 1	ST-1 complex ST-5 complex ST-1 complex	PorA_VR1 5-2 5-2 20 18-1	PorA_VR2 10 10 9 3	FetA_VR F1-5 F5-1 F3-1 F5-1	
27,521 (() () () () () () () () () (Order b Displa 1 records r (solate (M1027 120M 7891 6748 129E	25 ▼ records per page (eturned (1 - 25 displayed). Cl 1 2 3 4 5 6 eturned (1 - 25 displayed). Cl 1 3 4 5 6 eturned (1 - 25 displayed). Cl 1 3 4 5 6 eturned (1 - 25 displayed). Cl 1 3 4 5 6 eturned (1 - 25 displayed). Cl 1 3 4 5 6 etu	ick the hyperli Country USA Pakistan Finland Canada Germany	solate fid year 1937 i 1967 1975 i 1971 i 1964 i	elds () disease meningitis and septicaemia musaive (unspecified/other) meningitis and septicaemia	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Capsule grou A A A A A A	1 5 1	ST-1 complex ST-5 complex ST-1 complex ST-1 complex ST-1 complex	PorA_VR1 5-2 5-2 20 18-1 5-2	PorA_VR2 10 10 9 3 10	FetA_VR F1-5 F5-1 F3-1 F5-1 F3-6	
27,521 (() () () () () () () () () (Order b Displa 1 records r (solate (M1027 120M 7891 6748 129E 139M	25 ▼ records per page (eturned (1 - 25 displayed). Cl 1 2 3 4 5 6 B1; NIBSC 2803; Z1001 B35; NIBSC 2802; Z1001 B35; NIBSC 2784; Z1075 B54; NIBSC 2784; Z1075 B92; NIBSC 2784; Z1092 B99; NIBSC 2787; Z1099	ick the hyperli () () () () () () () () () ()	solate fid year 1937 i 1967 i 1975 i 1971 i 1964 i 1968	detailed information. disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	species) Neisseria meningitidis Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis	capsule grou A A A A A A A	1 5 1 1 1	Clonal complex ST-1 complex ST-5 complex ST-1 complex ST-1 complex ST-1 complex	PorA_VR1 5-2 5-2 20 18-1 5-2 5-2 5-2	PorA_VR2 10 10 9 3 10 10	FetA_VR F1-5 F5-1 F3-1 F3-1 F3-6 F5-1	
27,521 (() () () () () () () () () (Order b Displa 1 records r (solate (M1027 120M 7891 6748 129E	25 ▼ records per page (eturned (1 - 25 displayed). Cl 1 2 3 4 5 6 eturned (1 - 25 displayed). Cl 1 3 4 5 6 eturned (1 - 25 displayed). Cl 1 3 4 5 6 eturned (1 - 25 displayed). Cl 1 3 4 5 6 eturned (1 - 25 displayed). Cl 1 3 4 5 6 etu	ick the hyperli Country USA Pakistan Finland Canada Germany	solate fid year 1937 i 1967 i 1975 i 1971 i 1964 i 1968 1973 i	detailed information. disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other)	species) Neisseria meningitidis Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis) Neisseria meningitidis Neisseria meningitidis	Capsule grou A A A A A A	1 5 1 1 1 4	ST-1 complex ST-5 complex ST-1 complex ST-1 complex ST-1 complex	PorA_VR1 5-2 5-2 20 18-1 5-2	PorA_VR2 10 10 9 3 10	FetA_VR F1-5 F5-1 F3-1 F5-1 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.

	Sdb												
ome	e > Organisi	ms > Neisseria spp. > Neiss	eria isolates >	Search	n or browse database								
Se	arch o	or browse data	abase									Help 🔽	6] /
Ent	ter search cri	iteria or leave blank to brows	e all records. M	lodify fo	orm parameters to filter or e	enter a list of values.							
	Isolate pr	ovenance fields				Tagged sequence cour	its						
	id	× =	✓ Enter valu	0		ount of BACT000065 (r		v 2	2 +	(i)			
			Liner valu					. 4		0			
	Display/s	ort options				Action							
	Order b	by: id			✓ ascending ✓								
	Dicols	ay: 25 🗸 records per page	<u>~</u>			RESET SEA	ARCH						
	Displa	ister i records per page	0										
27,		returned (1 - 25 displayed). Cl 1 2 3 4 5 6	lick the hyperlin)									
(187 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6	lick the hyperlin	olate f	ields (į)				MLST		etyping anti		
((187 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases	lick the hyperlin	olate fi year	ields 🕞 disease		capsule group	o ST	MLST clonal complex	PorA_VR:	1 PorA_VR2	2 FetA_VR	
d 1	187 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allasca B1; NIBSC_2803; Z1001	Lick the hyperlin	olate f year 1937	ields (j) disease invasive (unspecified/other)) Neisseria meningitidis	capsule group A	ST	clonal complex	PorA_VR: 5-2	1 PorA_VR2 10	PetA_VR F1-5	
d 1 2	187 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	iick the hyperlii	olate f year 1937 1967	ields () disease invasive (unspecified/other) meningitis and septicaemia) Neisseria meningitidis Neisseria meningitidis	Capsule group A A	1	clonal complex ST-1 complex	PorA_VR: 5-2 5-2	1 PorA_VR2 10 10	2 FetA_VR F1-5 F5-1	
d 1 2 7	187 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 aliasos B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B34; NIBSC_2760; Z1054	lick the hyperlin () () () () () () () () () ()	olate f year 1937 1967 1975	ields () disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other)) Neisseria meningitidis Neisseria meningitidis) Neisseria meningitidis	Capsule group A A A	5 ST	clonal complex ST-1 complex ST-5 complex	PorA_VR: 5-2 5-2 20	1 PorA_VR2 10 10 9	PetA_VR F1-5	
d 1 2 7	187 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6 allases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	ick the hyperlin () () () () () () () () () ()	iolate fi year 1937 1967 1975 1964	ields () disease invasive (unspecified/other) meningitis and septicaemia) Neisseria meningitidis Neisseria meningitidis) Neisseria meningitidis	Capsule group A A	1 5 1	clonal complex ST-1 complex	PorA_VR: 5-2 5-2	1 PorA_VR2 10 10	FetA_VR F1-5 F5-1 F3-1	
d 1 2 7 .1 .3	187 records	allascs B1; NIBSC_2803; 21001 B55; NIBSC_2803; 21001 B55; NIBSC_2822; 21035 B54; NIBSC_2822; 21035 B54; NIBSC_2828; 21092	lick the hyperlin () () () () () () () () () ()	olate fi year 1937 1967 1975 1964 1968	ields () disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other)) Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Capsule group A A A A A	1 5 1 1	clonal complex ST-1 complex ST-5 complex ST-1 complex	PorA_VR: 5-2 5-2 20 5-2	1 PorA_VR2 10 10 9 10	2 FetA_VR F1-5 F5-1 F3-1 F3-6	
(«	187 records	allases allases B1; NIBSC_2803; Z1001 B55; NIBSC_2803; Z1001 B55; NIBSC_2760; Z1054 B92; NIBSC_2782; Z1092 B99; NIBSC_2795; Z1099	ick the hyperlin () () country USA Pakistan Finland Germany Philippines Ghana	olate fi year 1937 1967 1975 1964 1968 1973	ields () disease invasive (unspecified/other) meningitis and septicaemis invasive (unspecified/other) invasive (unspecified/other)	 Neisseria meningitidis 	Capsule group A A A A A	1 5 1 1 4	clonal complex ST-1 complex ST-5 complex ST-1 complex ST-1 complex	PorA_VR: 5-2 20 5-2 5-2 5-2	1 PorA_VR2 10 10 9 10 10 10	2 FetA_VR F1-5 F5-1 F3-1 F3-6 F5-1	
id 1 2 7 11 13 19	187 records isolate A4/M1027 120M 7891 129E 139M S3131	aliases aliases B1; NIBSC_2803; Z1001 B55; NIBSC_2803; Z1001 B55; NIBSC_2803; Z1001 B54; NIBSC_2760; Z105 B54; NIBSC_2760; Z1092 B99; NIBSC_2795; Z1099 B9213; NIBSC_2813; Z1213	lick the hyperlin () () () () () () () () () ()	olate fi year 1937 1967 1975 1964 1968 1973	ields () disease invasive (unspecified/other, meningitis and septicaemia invasive (unspecified/other, invasive (unspecified/other, invasive (unspecified/other,	 Neisseria meningitidis 	Capsule group A A A A A A A	1 5 1 1 4 5	clonal complex ST-1 complex ST-5 complex ST-1 complex ST-1 complex ST-4 complex	PorA_VR: 5-2 20 5-2 5-2 5-2 7	1 PorA_VR2 10 10 9 10 10 10 13-1	FetA_VR F1-5 F5-1 F3-1 F3-6 F5-1 F5-1	

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.

11.10.7 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.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	Help 🗹 🚺 🗡
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.	×
Isolate provenance fields Display/sort options	Modify form parameters
id • = • Enter value + () Order by: id Display: 25 • records per page () Action RESET SEARCH	Click to add or remove additional query terms: Provenance fields Secondary metadata Allele designations/scheme field values Allele designation counts
	Allele designation status Tagged sequence counts Tagged sequence status Attribute values list Filters

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.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	Help 🕜 🚺 🗡
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values. Isolate provenance fields Id V = V Enter value Display/sort options Order by: id V ascending V Action Display: 25 V records per page () RESET SEARCH	
1 record returned. Click the hyperlink for detailed information. Isolate fields () MLST Finetyping antigens Id Isolate aliases country year disease species capsule group ST clonal complex PorA_VR1 PorA_VR2 FetA_VR VR V	
Isolate fields () MLST Finetyping antigens Id Isolate allases country year disease species capsule group 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	
Isolate fields () MLST Finetyping antigens Id Isolate atlases country year disease species capsule group 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 Combinations Polymorphic sites Publications Sequence bin	
Isolate fields () MLST Finetyping antigens Id Isolate allases country year disease species capsule group 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	

Sequence tag flags

11.10.8 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.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	Help 🗹 🚯 🗡
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.	×
Isolate provenance fields Display/sort options	Modify form parameters
id v = V Enter value () Order by: id Display: 25 records per page ()	Click to add or remove additional query terms:
Action	Provenance fields
RESET	 Secondary metadata Allele designations/scheme field values
KESET SEARCH	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.

BIGSdb									
lome > Organisms > Neisseria spp. >	Neisseria isolates > Sea	rch or browse databas	se						
Search or browse d	atabase							Help	23/
Enter search criteria or leave blank to b	rowse all records. Modify	y form parameters to f	filter or enter a list of va	lues.					
Isolate provenance fields			— Attribute value	es list					
id 💙 =	✓ Enter value	+ (Field: id			~			
			1 2 3						
			3 4						
			5						
Display/part antions			Action						
 Display/sort options 			Action						
		× accondin							
Order by: id	age 🕥	✓ ascendin	ng 👻 RESET	SEARCH					
	age (i)	✓ ascendin	ng 👻 RESET	SEARCH					
Order by: id Display: 25 v records per p	U		ng v RESET	SEARCH					
Order by: id	U		ng × RESET	SEARCH					
Order by: id Display: 25 v records per p	for detailed information	ı.	ng v RESET	SEARCH		Final			
Order by: id Display: 25 v records per p	U	ı.	RESET		MLST		typing antic		
Order by: Id Display: 25 ▼ records per p 5 records returned. Click the hyperlinks Id isolate aliases 1 A4/M1027 B1; NIBSC_2803; Z1001	for detailed information Isolate fiel country year USA 1937 invasiv	n. .ds (j) disease ve (unspecified/other)	RESET species Neisseria meningitidis	capsule group A	MLST ST clonal comple	x PorA_VR1 5-2	PorA_VR2 10	FetA_VR F1-5	
Order by: id Display: 25 ✓ records per p 5 records returned. Click the hyperlinks 1 isolate allascs 1 A4/M1027 B1; NIBSC _2803; 21001 2 120M B55, NIBSC _2822; 2103	for detailed information Isolate fiel Country year USA 1937 invasiv 5 Pakistan 1967 menin	n. disease ve (unspecified/other) gitis and septicaemia	species Neisseria meningitidis Neisseria meningitidis	capsule group A A	MLST ST clonal comple 1 ST-1 complex	x PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR	
Order by: Id Display: 25 ▼ records per p 5 records returned. Click the hyperlinks Id isolate aliases 1 A4/M1027 B1; NIBSC_2803; Z1001	for detailed information Isolate fiel Country Year USA 1937 Invasiv Pakistan 1967 menin UK 2000 invasis	ı. disease ve (unspecified/other) gitis and septicaemia ve (unspecified/other)	RESET species Neisseria meningitidis	capsule group A A	MLST ST clonal comple	x PorA_VR1 5-2 5-2 19	PorA_VR2 10	FetA_VR F1-5	
Order by: Id Display: 25 ▼ records per p 5 records returned. Click the hyperlinks Id Isolate aliases 1 A4/M1027 B1; NIBSC_2803; Z1001 2 120M B35; NIBSC_2822; Z103 3 M00242905	for detailed information Isolate fiel (country year) USA 1937 invasiv 5 Pakistan 1967 menin UK 2000 invasiv 5 USA 1937 invasiv	n. disease ve (unspecified/other) gitis and septicaemia ve (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis	capsule group A A B A	MLST ST clonal complex 1 ST-1 complex 1099	x PorA_VR1 5-2 5-2 19	PorA_VR2 10 10	FetA_VR F1-5	
Order by: id Display: 25 ▼ records per p 5 records returned. Click the hyperlinks id isolate allases 1 A4/M1027 B1; NIBSC_2803; Z1001 2 120M B55; NIBSC_2822; Z103 3 M00242905 4 M1027 B43; NIBSC_3076; Z104	for detailed information Isolate fiel (country year) USA 1937 invasiv 5 Pakistan 1967 menin UK 2000 invasiv 5 USA 1937 invasiv	n. disease ve (unspecified/other) gitis and septicaemia ve (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group A A B A	MLST ST clonal comples 1 ST-1 comples 1099 4 ST-4 comples	x PorA_VR1 5-2 5-2 19	PorA_VR2 10 10 15	FetA_VR F1-5	
Order by: id Display: 25 ▼ records per p 5 records returned. Click the hyperlinks id isolate allases 1 A4/M1027 B1; NIBSC_2803; Z1001 2 120M B55; NIBSC_2822; Z103 3 M00242905 4 M1027 B43; NIBSC_3076; Z104	for detailed information Isolate fiel (country year) USA 1937 invasiv 5 Pakistan 1967 menin UK 2000 invasiv 5 USA 1937 invasiv	n. disease ve (unspecified/other) gitis and septicaemia ve (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group A A B A	MLST ST clonal comples 1 ST-1 comples 1099 4 ST-4 comples	x PorA_VR1 5-2 5-2 19	PorA_VR2 10 10 15	FetA_VR F1-5	
Order by: id Display: 25 ✓ records per p 5 records returned. Click the hyperlinks 1 A4/M1027 81; NIBSC_2803; Z1001 2 120M 835; NIBSC_2822; Z103 3 M00242905 4 M1027 843; NIBSC_3076; Z104 5 M00240227	for detailed information Isolate fiel Country year USA 1937 invasi 9 Pakistan 1967 menin UK 2000 invasiv UK 2000 invasiv	n. disease ve (unspecified/other) gitis and septicaemia ve (unspecified/other) ve (unspecified/other) ve (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group A A B B B B	MLST ST clonal comple 1 ST-1 complex 1099 4 ST-4 complex	x PorA_VR1 5-2 5-2 19	PorA_VR2 10 10 15	FetA_VR F1-5	
Order by: Id Display: 25 ▼ records per p 5 records returned. Click the hyperlinks 1 A4/M1027 B1; NIBSC_2803; Z1001 2 120M B35; NIBSC_2803; Z1003 3 M00242905 4 M1027 B43; NIBSC_3076; Z104 5 M00240227 Analysis tools	for detailed information Isolate fiel Country Year USA 1937 invasis Pakistan 1967 menin UK 2000 invasis UK 2000 invasis UK 2000 invasis UK 2000 invasis	n. disease ve (unspecified/other) gitis and septicaemia ve (unspecified/other) ve (unspecified/other) ve (unspecified/other) Hymorphic sites Put	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group A A B B B : in	MLST ST clonal comple 1 ST-1 complex 1099 4 ST-4 complex	x PorA_VR1 5-2 5-2 19	PorA_VR2 10 10 15	FetA_VR F1-5	
Order by: Id Display: 25 ▼ records per p 5 records returned. Click the hyperlinks 1 A4/M1027 B1; NIBSC_2803; Z1001 2 120M B35; NIBSC_2803; Z1003 3 M00242905 4 M1027 B43; NIBSC_3076; Z104 5 M00240227 Analysis tools € Breakdown: Fields Two Fiel	for detailed information Isolate fiel Country Vear USA 1937 invasik Pakistan 1967 menin UK 2000 invasik UK 2000 invasik UK 2000 invasik UK 2000 invasik UK 2000 invasik d Combinations Po Gene Presence Get	n. disease ve (unspecified/other) gitis and septicaemia ve (unspecified/other) ve (unspecified/other) ve (unspecified/other) Hymorphic sites Put	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group A A B B B : in	MLST ST clonal comple 1 ST-1 complex 1099 4 ST-4 complex	x PorA_VR1 5-2 5-2 19	PorA_VR2 10 10 15	FetA_VR F1-5	

11.10.9 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.

BIGSdb			
Home > Organisms > Neisseria spp. >	Neisseria isolates > Search or browse database	e	
Search or browse c	latabase		Help 🗹 🚺 🗡
Enter search criteria or leave blank to	browse all records. Modify form parameters to fi	ilter or enter a list of values.	×
Isolate provenance fields		Display/sort options	Modify form parameters
id v = Action RESET SEARCH	 ✓ Enter value (1) 	Order by: id Display: 25 ✓ records per page ()	Click to add or remove additional query terms: Provenance fields Secondary metadata Allele designations/scheme field values Allele designation counts Allele designation status Tagged sequence counts Tagged sequence status Altribute values list Filters Filters

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.
- **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*.
- Old record versions Checkbox which, if selected, will include all record versions in a query.

11.10.10 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.

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.

BIGSdb			
Home > Organisms > Neisseria spp. > Neisseria isolates			
Neisseria isolates database			
The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria sy database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and conseque sample.			
Query database		→ LOG IN	
Search database			
Browse, search by any criteria, or enter list of attributes.		PROJECTS	+
	1	Z EXPORT	+
Search by combinations of loci This can include partial matches to find related isolates.			
			+
			+

If multiple schemes are defined, you may have to select the scheme you wish to query in the 'Schemes' dropdown box and click 'Select'.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search by locus combinations	
Search by locus combinations	Help 📝 🚺
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	
Project: Search: Exact or nearest match •	
□ Include old record versions Order by: id v ascending v Display: 25 v records per page ①	
Action	
RESET SEARCH	

Enter the combination of alleles that you want to query for. Fields can be left blank.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search by locus combinations	
Search by locus combinations	Help 🕜 🕕 🚺
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 database	
abcZ adk aroE fumC gdh pdhC pgm ST: Autofill 2 3 4 3 8 4 6 6	
Filters Options	
Project: Search: Exact or nearest match •	
Include old record versions Display/sort options	
Order by:	
id value ascending v Display: 25 v records per page ()	
Action	
RESET SEARCH	

Alternatively, for scheme profiles, you can enter a primary key value (e.g. ST) and select 'Autofill' to automatically fill in the associated profile.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search by locus combinations	
Search by locus combinations	Help 🕜 🚺 🚼
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: 44 Autofill 9 6 9 9 6 9	
Filters Options	
Project: Search: Exact or nearest match •	
Include old record versions Display/sort options	
Order by: id	
Display: 25 v records per page ()	
Action	
RESET SEARCH	

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.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search by locus combinations	
Search by locus combinations	Help 🕜 3
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 database abcZ adk aroE fumC gdh pdhC pgm ST: 44 Autofill 9 6 9 9 6 9	
Filters Options	
Project: Search: Exact or nearest match •	
Include old record versions Display/sort options	
Order by:	
id	
Action	
RESET SEARCH	

Click 'Search'.

	db										
lome	> Organism	s > Neisseria spp. > 1	Neisseria isolates	> Search	by locus combinations						
											Help 🕜 🔒 🔡
Sea	arch b	y locus coi	mbinatic	ons							
		/									
Sch	hemes										
Plea	se select the	scheme you would lik	e to query:								
		Scheme you would like									
MLS	ST		~	Select							
	Please ente	er your allelic profile	below. Blank lo	ci will be	ignored.	— — Autofill profile	by searching	remote database			
	abcZ 9	adk aroE 6 9	fumC 9	gdh 9	pdhC pgm 6 9	ST: 44	4	Autofill			
		0 9	9	9							
	Filters				Options						
	Project			~ (i)	Search: Exact	t or nearest match 🗸					
0	Include old	d record versions			 Display/sort optic 	ons					
					Order by:						
					id		✓ as	cending 🗸			
					Display: 25	 records per page (i) 					
	Action										
	DECET	CEADCH.									
	RESET	SEARCH									
5											
	t matches fo										
		ound (7 loci). rned (1 - 25 displayed).	Click the hyperlir	nks for det	ailed information.						
195	records retur	rned (1 - 25 displayed).		~	ailed information.						
	records retur			iks for det	ailed information.						
195	records retur	rned (1 - 25 displayed).	6 🕥 🤅					MIST	Finety	ping antig	nens
195	records retur	rned (1 - 25 displayed).	6 🕥 🤅	~		species	capsule group	MLST ST clonal complex	Finety PorA_VR1	ping antig PorA_VR2	
195 《	records retur	rned (1 - 25 displayed). 1 2 3 4 5	6 () () country Germany	solate fiel year 1999	ds (j) disease carrier	Neisseria meningitidis	В	ST clonal complex 44 ST-41/44 complex			
195 《 id 41 70	records retur	nned (1 - 25 displayed). 1 2 3 4 5 aliases	6 () () country Germany Germany	solate fiel year 1999 1999	ds (j) disease carrier carrier	Neisseria meningitidis Neisseria meningitidis	B NG	STclonal complex44ST-41/44 complex44ST-41/44 complex	PorA_VR1	PorA_VR2	FetA_VR
195 《 id 41	records retur	rned (1 - 25 displayed). 1 2 3 4 5	6 () () country Germany	solate fiel year 1999 1999 1988	ds (j) disease carrier carrier carrier	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B NG B	ST clonal complex 44 ST-41/44 complex			
195 (() id 41 70 427	isolate 19 38 NG E30	nned (1 - 25 displayed). 1 2 3 4 5 aliases	6 () () country Germany Germany Norway	solate fiel year 1999 1999 1988 1999 inv	ds (j) disease carrier carrier	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B NG B B	ST clonal complex 44 ST-41/44 complex 44 ST-41/44 complex 44 ST-41/44 complex 44 ST-41/44 complex	PorA_VR1	PorA_VR2	FetA_VR
195 (() id 41 70 427 774 792 975	isolate 19 38 NG E30 99 182 99-132 0069/93	1 25 displayed). 1 2 3 4 5 allases NIBSC_2804; Z4692	6 () () Germany Germany Norway Canada Canada Czech Republic	solate fiel year 1999 1999 1988 1999 inv 1999 inv 1993	ds () disease carrier carrier carrier asive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B NG B B B B B	ST clonal complex 44 ST-41/44 complex 45 ST-41/44 complex 45 ST-41/44 complex 45 ST-41/44 complex 45 ST-41/44 complex	PorA_VR1	PorA_VR2 16 14-4	FetA_VR
195 id 41 70 427 774 792 975 1097	isolate 19 38 NG E30 99 182 99-132 0069/93 3552	1 25 displayed). 1 2 3 4 5 allases NIBSC_2804; Z4692	6 Country Germany Germany Norway Canada Canada Czech Republic The Netherlands	solate fiel year 1999 1999 1988 1999 inv 1999 inv 1993 1975	ds () disease carrier carrier asive (unspecified/other) asive (unspecified/other) carrier	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B NG B B B B B B	ST clonal complex 44 ST-41/44 complex 45 ST-41/44 complex 44 ST-41/44 complex 44 ST-41/44 complex 45 ST-41/44 complex 44 ST-41/44 complex	21 22	PorA_VR2 16 14-4 4	FetA_VR F1-7 F1-7
195 (« 41 70 427 774 792 975 1097 1603	isolate 19 38 NG E30 99 182 99-132 0069/93 3532 0213/93	1 25 displayed). 1 2 3 4 5 allases NIBSC_2804; Z4692	6 Country Germany Germany Norway Canada Canada Cach Republic The Netherlands Zeech Republic	solate fiel year 1999 1999 1988 1999 inv 1999 inv 1993 1975 1993	ds () disease carrier carrier carrier asive (unspecified/other) carrier carrier	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B NG B B B B B B	ST clonal complex 44 ST-41/44 complex 45 ST-41/44 complex 45 ST-41/44 complex 44 ST-41/44 complex	PorA_VR1 21 22 22	PorA_VR2 16 14-4 4 14-4	FetA_VR F1-7 F1-7 F1-19
195 id 41 70 427 774 792 975 1097	isolate 19 38 NG E30 99 182 99-132 0069/93 3552	1 25 displayed). 1 2 3 4 5 allases NIBSC_2804; Z4692	6 Country Germany Germany Norway Canada Canada Czech Republic The Netherlands	solate fiel year 1999 1988 1999 inv 1999 inv 1993 1975 1993 1993	ds () disease carrier carrier asive (unspecified/other) asive (unspecified/other) carrier	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	B NG B B B B B B B B B B	ST clonal complex 44 ST-41/44 complex 45 ST-41/44 complex 44 ST-41/44 complex 44 ST-41/44 complex 45 ST-41/44 complex 44 ST-41/44 complex	21 22	PorA_VR2 16 14-4 4	FetA_VR F1-7 F1-7

11.11 Bookmarking an isolate query

Once you have made an isolate database query, you can bookmark it so that it can be repeated in the future. You need to have an account and be logged in to the database to be able to bookmark.

Following a query, there will be a 'Bookmark query' section in the results header section. Enter a name for the bookmark (a default name based on the date will be shown) and click 'Add bookmark'.

Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.	
Isolate provenance fields Filters	
Combine with: AND > Publication: Select options Image: Combine with: Select optine with: Select options Image: Combine	
Add fitter: Add Display/sort options Order by: id Display: 25 v records per page () Action RESET SEARCH	
2,302 records returned (1 - 25 displayed). Click the hyperlinks for detailed information. Your projects Bookmark query 2020-07-28:1 Add bookmark	
Isolate fields 🕜 MLST Finetyping antigens	
id isolate aliases country year disease species capsule group ST clonal complex POrA_VR1 PorA_VR2 FetA_VR	
30154 MI4 240001 UK [England] 2014 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex 5 2 F1-1	
30155 M14 240002 UK [England] 2014 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex 5 2 F1-1	
30156 M14 240007 UK [England] 2014 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex 5 2 F1-1 30157 M14 240013 UK [England] 2014 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex 5 2 F1-1	
30157 M14 240013 UK [England] 2014 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex 5 2 F1-1 30158 M14 240019 UK [England] 2014 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex 5 2 F1-1	
50159 M14 240012 UK [England] 2014 invasive (unspecified/other) Neisseria meningituitis W 8621 ST-11 complex 5 2 F1-1	
30160 M12 420022 UK [England] 2014 invasive (unspecified/other) Neisseria meningituitis W 11 ST-11 tomplex 5 2 F1-1	
30161 H14 240029 UK [England] 2014 invasive (unspecting/other) Neisseria meningitidis W 11 ST-11 complex 5 2 F1-1	
30162 M14 240031 UK [England] 2014 invasive (unspecified/other) Neisseria meningitidis W 11 ST-11 complex 5 2 F1-1	

A new top-right tab will appear when you have saved bookmarks. You can click on this to easily access all saved bookmarks.

BIGSdb						
Home > Organisms	> Neisseria spp. > N	leisseria isolates > Search or	browse database			
Search or	browse d	atabase			🚹 Help 🗹	3 🕄 🗡 🔲
Enter search criteria	a or leave blank to b	rowse all records. Modify form	parameters to filter or	enter a list of values.		×
Isolate prove	nance fields			Display/sort options		Bookmarks
id	♥ =	 ✓ Enter value 	+ (j	Order by: id Display: 25 🕶 records per page 👔	✓ ascent	• 2020-07-28:1
RESET	SEARCH					Manage bookmarks

You can also access your bookmarks from the main contents page. A link will appear in the query section once you have saved a bookmark.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria isolates		
Neisseria isolates database		
The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria speci database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently sample.		1
Query database	€ LOG OUT	+
Search database	1 SUBMISSIONS	
Browse, search by any criteria, or enter list of attributes.	PRIVATE DATA	
Search by combinations of loci	PROJECTS	+
This can include partial matches to find related isolates.	Z EXPORT	+
	MALYSIS	
Bookmarks	CUSTOMISE	+
Retrieve dataset from bookmarked queries.	INFORMATION	+
	🚔 JOBS 📵	

This will take you to a page where you can manage your bookmarks.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Bookmarks	
Bookmarks	Help 🗹
Please note that only you will be able to access a query defined by a bookmark if it is shown as locked. Click the padlock icon to make it publicly shareable - you can then copy the URL from the 'Run query' field and share this with colleagues. Delete Vare Database configuration Created Share Run query X 2020-07-28:1 pubmlst_neisseria_isolates 2020-07-28 Q	

You can go to a bookmarked search by clicking on the 'Run query' icon. By default, a bookmark can only be used by the logged-in user who created it. This is for privacy reasons to prevent other users from finding out what terms are being used for a search. If, however, you wish to share the URL to the query, you can make it shareable by clicking the padlock icon.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Bookmarks	
Bookmarks	Help 🕜
Please note that only you will be able to access a query defined by a bookmark if it is shown as locked. Click the padlock icon to make it publicly shareable - you can then copy the URL from the 'Run query' field and share this with colleagues. Delete Name Database configuration Created Share Run query 2020-07-28:1 pubmlst_neisseria_isolates 2020-07-28	

The icon will change to a green open padlock. You can right-click on the 'Run query' link to copy the URL if you wish to share it with others.



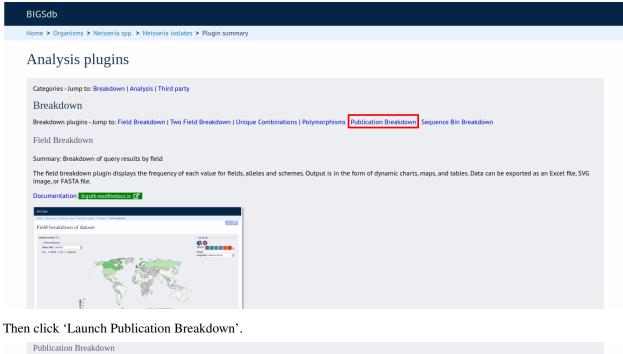
Bookmarks can be deleted by clicking on the delete icon.

11.12 Retrieving isolates by linked publication

Click 'Analysis' on the contents page.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria isolates		
Neisseria isolates database		
The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria sp database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequer sample.		
Query database	⊕ LOG IN	
Search database		
Browse, search by any criteria, or enter list of attributes.	E PROJECTS	+
	C EXPORT	+
Search by combinations of loci This can include partial matches to find related isolates.	MALYSIS	
		+
	1 INFORMATION	+

Click 'Publication breakdown'



ummary: Breakdown of query results by publication
animary, Breakdown of query results by publication
his plugin shows all publications linked to isolates in a query dataset or within the whole database. The results can be filtered by author or year. The output includes full citation details and a link t isplay all isolates linked to any listed publication.
bocumentation bigstbreatthedocsio G
94.00° bit of parts hundrings historia shares higher hikitates
Publication breakdown of dataset
- Pernamying Budy Attention Attentio
31 meter energi - 31 égenete la la ingente la catacitationa. ⊕ () 1 : 0 ()
Name and Sector Control Contro
2012/R 2012 Mark Mark (2014) Cally A, Mark (1) Mark (2014) Kears (3) Lill A. Jaywa per yanaking participation and an animative and antipole and and and an animative and antipole and and and an animative a
BIOIN HILMARK Agana, Add. Swol, Lang, Dav. Link, K. Kataka, asson apply a protocol and and and an anti-table method of a strange of
Launch 'Publication Breakdown'
equence Bin Breakdown
ummary: Breakdown of sequence bin contig properties
he 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 each isolate record. Using this attei netric against a cgMLST scheme can be a good indicator of genome quality. Values for number of contigs, total sequence length, mean contig length and contig length distribution are charted.
Jocumentation bigsdb.readthedocs.io 🗗
tón I

A list of publications linked by isolates within the database will be displayed.

BIGSdb				
lome > Orga	anism	ns > Neisseria spp. > Neisseria isolates > Plugins > Publications		
Public	ati	ion breakdown of dataset		Help 🔽
	utho	y by Display Action r: All authors V Order by: number of isolates V descending SUBMIT SUBMIT		
138 records	s retu	rned (1 - 25 displayed). Click the hyperlinks for detailed information.		
(() ()	s retu) Year	Inred (1 - 25 displayed). Click the hyperlinks for detailed information. 1 2 3 4 5 6 () (>) Clicktion	Title	Isolates in database
Image: Weight of the second	Year	1 2 3 4 5 6 🔊 🔊		isolates in database 3705 isolates
WibMed Y id Y 321635580 20	Year 2020	1 2 3 4 5 6 (>) Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial Resistance. Typing complex meningococcal vaccines to understand diversity and population	database
(*) (*) PubMed id ¥ 32163580 20 30687793 20	Year 2020 2018	1 2 3 4 5 6 Solution Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of print]: Rodrigues CMC, Chan H, Vipond C, Jolley K, Harrison OB, Wheeler J, Whiting G, Feavers IM, Maiden MCJ (2018) Wellcome	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial Resistance. Typing complex meningococcal vaccines to	database 3705 isolates

These can be filtered by author and/or year, and the sort order changed.

BIGSdb				
lome > O	rganisn	ms > Neisseria spp. > Neisseria isolates > Plugins > Publications		
Publi	cat	ion breakdown of dataset		Help 🔽
Filte		rry by Display Action cr: Jolley KA ▼ Order by: number of isolates ▼ descending ▼ SUBMIT ar: All years ▼ Display: 25 ▼ records per page		
19 record PubMed id	ds retur Year	rned. Click the hyperlinks for detailed information. Citation	Title	Isolates in database
PubMed id	Year	Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of print]:	Title Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial Resistance.	
PubMed id 32163580	Year 2020 2020	Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of print]: 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, Kesanopoulos K, Kriz P, Lametrsen L, Levenet I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G,	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial	database
PubMed id 32163580 17517841	Year 2020 2007	Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of print]: 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, Koesanopoulos K, Kriz P, Lambertsen L, Levene L, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanetill 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 Natl Acad Sci U S A 105 : 15082-7	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial Resistance. Target gene sequencing to characterize the penicillin 6 susceptibility of Neisseria	database 3705 isolates
PubMed id 32163580 17517841 18815379	 Year 2020 2007 2008 2015 	Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of print]: 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, Kchrid A, Kesanopoulos K, Kriz P, Lambertsen L, Levene I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanetill 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 Natl Acad Sci U S A 105 : 15082-7 Hill DM, Lucidarme J, Gray SJ, Newbold LS, Ure R, Brehony C, Harrison OB, Bray JE, Jolley KA, Bratcher HB, Parkhill J, Tang CM, Borrow R, Maiden MC (2015) Lancet Infect Dis 15 : 1420-8	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial Resistance. Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis. Role of selection in the emergence of lineages and the evolution of virulence in	database 3705 isolates 1666 isolates

To display the isolate records for any of the displayed publications, click the button to the right of the citation.

BIGSdb				
lome > O	rganisi	ns > Neisseria spp. > Neisseria isolates > Plugins > Publications		
Publi	cat	ion breakdown of dataset		Help 🗹
——Filte		ry by Display Action r: Jolley KA Order by: number of isolates V descending V r: All years Display: 25 V records per page SUBMIT		
	ls retu	ned. Click the hyperlinks for detailed information.		Isolatos in
19 record PubMed id	ds retu Year	ned. Click the hyperlinks for detailed information. Citation	Title	lsolates in database
PubMed id	Year			
PubMed id 32163580	Year 2020 2007	Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial	database
PubMed id 32163580 17517841	Year 2020 2007	Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of print]: Taha MK, Väzquez 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, Hoffman S, Jolley KA, Kadlubowski M, Kersanopoulos K, Kriz P, Lambertsen L, Levenet I, Musilek M, Paragi M, Saguer A, Skoczynska A, Stefanetill P, Thulin S, Izanakaki G,	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial Resistance. Target gene sequencing to characterize the penicillin G susceptibility of Neisseria	database 3705 isolates
PubMed id 32163580 17517841 18815379	Year 2020 2007 2008	Citation Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of print]: 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, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother 51 : 2784-92	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial Resistance. Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis. Role of selection in the emergence of lineages and the evolution of virulence in	database 3705 isolates 1666 isolates

The abstract of the paper will be displayed (if available), along with all isolates linked to it.

)											
Home > C	Organisms > N	leisseria spp. > Neisseria isolates	> Publication	dataset								
Publi	ication	dataset										8
Citatio	n query (I	PubMed id: 32163580)										
Clusterir findings resistand	ng with cgMLS 6 have implicati 1 have than others.	Population Genomics: Use of the T identified globally distributed, pr ions for the emergence of antimicr	ersistent, gonoc obial resistanco	coccal line e in gonoc	ages improving under cocci and how this is a	standing of the population	biology of gonoc					
(()		1 (1 - 25 displayed). Click the hyper 2 3 4 5 6	links for detaile	ed informa	ition.							
.,/05 re			<u>»</u>	ed informa	tion.				MLST	Finet	yping anti	gens
id			<u>»</u>		tion. disease	species	capsule group		MLST		yping anti	
(() id 2855	3 1 isolate FA1090	2 3 4 5 6 () aliases BennettTree06; NC_002946	Isolate 1 country USA	fields (j) year 1980		Neisseria gonorrhoeae		ST c 1899		PorA_VR1 18-10	PorA_VR2 43	FetA_VR F1-27
id 2855 13685	3 1 isolate FA1090 NCCP11945	2 3 4 5 6 (2) alises BennetTree06; NC_002946 BennetTree04; NC_011035	Isolate 1 country USA South Korea	fields (j) year 1980 2005		Neisseria gonorrhoeae Neisseria gonorrhoeae		ST c 1899 1901		PorA_VR1 18-10 18-10	PorA_VR2 43 43	FetA_VR F1-27 F3-14
id 2855 13685 1 14729	isolate FA1090 NCCP11945 MS-11	2 3 4 5 6 () aliases BennettTree06; NC_002946	Isolate 1 country USA South Korea USA	fields (j) year 1980		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae		ST c 1899 1901 1959		PorA_VR1 18-10 18-10 18-10 18-10	PorA_VR2 43 43 43	FetA_VR F1-27 F3-14 F4-7
id 2855 13685 14729 15698	3 1 isolate FA1090 NCCP11945 MS-11 DGI2	2 3 4 5 6 () aliases BennettTree06; NC_002946 BennettTree04; NC_011035 BennettTree05; NC_022240	Isolate 1 Country USA South Korea USA USA	fields () year 1980 2005		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae		ST c 1899 1901 5959 8421		PorA_VR1 18-10 18-10 18-10 18-10 18-10	PorA_VR2 43 43 43 43 43	FetA_VR F1-27 F3-14 F4-7 F3-14
id 2855 13685 14729 15698 15935	isolate FA1090 NCCP11945 MS-11 DGI2 PID24	2 3 4 5 6 () () aliases BennettTree06; NC_002946 BennettTree04; NC_011055 BennettTree05; NC_022240 BennettTree01	Isolate 1 Country USA South Korea USA USA USA	fields () year 1980 2005		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae		ST c 1899 1901 1959 8421 8418		PorA_VR1 18-10 18-10 18-10 18-10 18-10 18-10	PorA_VR2 43 43 43 43 43 43	FetA_VR F1-27 F3-14 F4-7 F3-14 F1-26
id 2855 13685 14729 15698 15935 15936	isolate FA1090 NCCP11945 MS-11 DGI2 PID24 DGI18	2 3 4 5 6 () aliases BennettTree06; NC_002946 BennettTree04; NC_011035 BennettTree05; NC_022240	Isolate 1 Country USA South Korea USA USA USA	fields () year 1980 2005 1960		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae		ST c 1899 1901 1901 1959 3421 18418 3418 18418		PorA_VR1 18-10 18-10 18-10 18-10 18-10 18-10	PorA_VR2 43 43 43 43 43 43 43 43	FetA_VR F1-27 F3-14 F4-7 F3-14 F1-26 F1-26
id 2855 13685 14729 15698 15935 15936	isolate FA1090 NCCP11945 MS-11 DGI2 PID24	2 3 4 5 6 () () aliases BennettTree06; NC_002946 BennettTree04; NC_011055 BennettTree05; NC_022240 BennettTree01	Isolate 1 Country USA South Korea USA USA USA	fields (j) year 1980 2005		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae		ST c 1899 1901 1959 8421 8418		PorA_VR1 18-10 18-10 18-10 18-10 18-10 18-10	PorA_VR2 43 43 43 43 43 43	FetA_VR F1-27 F3-14 F4-7 F3-14 F1-26
(() id 2855	isolate FA1090 NCCP11945 MS-11 DGI2 PID24 DGI18 TCDC-	2 3 4 5 6 () () aliases BennettTree06; NC_002946 BennettTree04; NC_011055 BennettTree05; NC_022240 BennettTree01	Isolate 1 Country USA South Korea USA USA USA	fields () year 1980 2005 1960		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae		ST c 1899 1901 1901 1959 3421 18418 3418 18418		PorA_VR1 18-10 18-10 18-10 18-10 18-10 18-10	PorA_VR2 43 43 43 43 43 43 43 43	FetA_VR F1-27 F3-14 F4-7 F3-14 F1-26 F1-26

11.13 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 'Customise' link on the database contents page. Most of the available options are visible for isolate databases, whereas sequence definition databases have fewer available.

BIGSdb						
Home > Organisms > Neisseria spp. > Neisseria isolates						
<i>Neisseria</i> i	isolates database					
	.ST database contains data for a collection of isolates that represent the total known diversity of Neisseria spe least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequen			n		
	Query database		→J LOG IN			
Sea	arch database					
Bro	owse, search by any criteria, or enter list of attributes.			+		
			EXPORT	+		
	a <mark>rch by combinations of loci</mark> is can include partial matches to find related isolates.		MALYSIS			
			¢ CUSTOMISE	-		
			General options Locus display Scheme display Scheme field display			
				+		

11.13.1 General options

The general options tab is displayed by default. If another tab is being shown, click the 'General options' header.

lGSdb
ome > Organisms > Neisseria spp. > Neisseria isolates > General options
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 isolates) 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 v records per page. Page bar position: top and bottom v Display 100 v inucleotides of flanking sequence (where available). Display locus aliases if set. C Enable tooltips (beginner's mode). SET OPTIONS
Main results table - display options
Main results table - provenance field selection
Main results table - secondary metadata selection
Isolate record 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 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.

11.13.2 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.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > General options
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 isolates) 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 v records per page. Page bar position: [top and bottom v] Display [100 v] nucleotides per line in sequence alignments. Display [100 v] nucleotides of flanking sequence (where available). □Display locus aliases if set. I Enable tooltips (beginner's mode). SET OPTIONS
Main results table - display options
Main results table - provenance field selection
Main results table - secondary metadata selection
Isolate record 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
'main results table' tab will scroll up.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > General options	
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 isolates) 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 - display options	
Differentiate provisional allele designations. Display information about sequence bin records tagged with locus information (tooltip). Display sequence bin size. Display contig count. Display publications. SET OPTIONS	
Main results table - provenance field selection	
Main results table - secondary metadata selection	
Isolate record display	
Query filters	
Reset Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.	

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.

11.13.3 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.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > General options
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 isolates) 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 v records per page. Page bar position: top and bottom v Display 100 v nucleotides per line in sequence alignments. Display 100 v nucleotides of flanking sequence (where available). Display locus aliases if set. Enable tooltips (beginner's mode). SET OPTIONS
Main results table - display options
→ Main results table - provenance field selection
→ Main results table - secondary metadata selection
→ Isolate record 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.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > General options
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 isolates) 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 - display options
Main results table - provenance field selection
Main results table - secondary metadata selection
 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). SET OPTIONS
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.

11.13.4 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.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > General options
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 isolates) 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 v records per page. Page bar position: [top and bottom v] Display 100 v nucleotides of flanking sequence alignments. Display 100 v nucleotides of flanking sequence (where available). Display locus aliases if set. Zenable tooltips (beginner's mode). SET OPTIONS
Main results table - display options
Main results table - provenance field selection
Main results table - secondary metadata selection
Isolate record display
Query filters
Reset Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

The 'Provenance field display' tab will scroll up.

BIGSdb			
Home > Organisms > Neisseria spp. > Neisseria	isolates > General options		
Set database options			Help 🔽
Here you can set options for your use of the web when you next go to a query page, try refreshing			ria isolates) only. If some of the options don't appear to set
General options			
• Main results table - display options			
✓ Main results table - provenance field selection			
Select the isolate provenance fields that y performing a locus, scheme or scheme field aliases strain_designation continent region year date_sampled soyear_sampled date_received non_culture epidemiological_year age_yr age_range		esults table following a query. Settings for displa omize' option. penicillin_range amoxicillin sulphonamide ceftriaxone ceftriaxone_range chloramphenicol_range cefotaxime_range frifampicin_range ciprofloxacin ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range ciprofloxacin_range	aying locus and scheme data can be made by azithromycin spectinomycin bioproject_accession ENA_run_accession Finvate_project comments sender curator date_entered datestamp
ALL NONE DEFAULT	SET OPTIONS		
• Main results table - secondary metadata selection	on		
▶ Isolate record display			
• Query filters			
Reset Click the reset button to remove all user settings RESET ALL TO DEFAULTS	for this database - this includes locus a	and 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'.

11.13.5 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.

		Help
Set database options		_
	ebsite. Options are remembered between sessions and affect th ng the page (Shift + Refresh) as some pages are cached by your	e current database (Neisseria isolates) only. If some of the options don't appear to se browser.
✓ General options		
Display 25 ✓ records per page. Page bar position: top and bottom ✓ Display 100 ✓ nucleotides per line in: Display 100 ✓ nucleotides of flanki		
 Display locus aliases if set. Enable tooltips (beginner's mode). 	······································	
SET OPTIONS		
• Main results table - display options		
• Main results table - provenance field selectio	n	
• Main results table - secondary metadata selec	ction	
• Isolate record display		
• Query filters		
Reset Click the reset button to remove all user settin RESET ALL TO DEFAULTS	ngs for this database - this includes locus and scheme field pref	rrences.
Click the reset button to remove all user settin		vrences.
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scre		rrences.
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scru Query filtes	oll up.	
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scru Query filtes	oll up.	erences.
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scru Query filtes Select the fields for which you would like d interface. isolate	oll up. Iropdown lists containing known values on which to filter qu	ery results. These will be available in the filters section of the query
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like of interface. Solate Country Continent	oll up. fropdown lists containing known values on which to filter qu sero_subtype 	ery results. These will be available in the filters section of the query private_project comments sender
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scru Query filters Select the fields for which you would like d interface. isolate ountry continent iregion	oll up. fropdown lists containing known values on which to filter quarter of the sero subtype Sero_subtype ET_no pericillin	ery results. These will be available in the filters section of the query private_project comments cender curator
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface. isolate country continent region year	oll up. dropdown lists containing known values on which to filter qu sero_subtype ET_no pericillin pericillin_range	ery results. These will be available in the filters section of the query private_project comments sender curator date_entered
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scru Query filtes Select the fields for which you would like d interface. isolate country continent region year date_sampled isoyear_sampled	oll up. dropdown lists containing known values on which to filter qu sero_subtype ET_no penicillin penicillin penicillin uphonamide	ery results. These will be available in the filters section of the query private_project comments sender curator date_entered date_stamp MLST profile completion
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface. isolate continent contin	oll up. tropdown lists containing known values on which to filter qu serotype Sero_subtype ET_no penicillin penicillin suphonamide ceftriaxone	ery results. These will be available in the filters section of the query private_project comments sender curator date_entered datestamp MLST profile completion Bexsero Antigen Sequence Typing (BAST) profile
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scrue Query filters Query filters Select the fields for which you would like of interface. Select the fields for which you would like of interface. Select the fields for which you would like of contry Continent Contry Contry Continent Contry Continent Contry Continent Contry Contry Continent Contry Contrinet Contry Contr	oll up. fropdown lists containing known values on which to filter quarter of the sero_subtype Sero_subtype ET_no Penicillin_range anoxicillin Suphonamide Ceftriaxone Ceftriaxone Ceftriaxone_range	ery results. These will be available in the filters section of the query private_project comments sender curator date_entered datestamp MLST profile completion CBessero Artigen Sequence Typing (BAST) profile completion
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface. isolate country continent egion year_sampled isoyear_sampled lisoyear_sampled onon_culture	oll up. dropdown lists containing known values on which to filter qu sero_subtype ET_no penicillin penicillin_range amoxicillin suphonamide ceftriaxone ceftriaxone_range chloramphenicol	ery results. These will be available in the filters section of the query private_project comments sender curator date_entered datestamp MLST profile completion Bexsero Antigen Sequence Typing (BAST) profile
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scru Query filters Select the fields for which you would like d interface. Select the fields for which you would like d interf	oll up.	hery results. These will be available in the filters section of the query private_project comments sender curator date_entered date_stamp MLST profile completion Bexsero Antigen Sequence Typing (BAST) profile completion Conjugative Plasmid profile completion GC ONU peptide typing profile completion GC Combustion
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface. isolate country continent geage isovear_sampled isovear_sampled isovear_sampled isovear_asampled isovear_	oll up. tropdown lists containing known values on which to filter qu	tery results. These will be available in the filters section of the query private_project comments sender curator date_entered datestamp MLST profile completion Conjugative Plasmid profile completion Cryptic Plasmid profile completion G OMV peptide typing profile completion N. gonorthoeae cgMLST v1.0 profile completion N. gonorthoeaee dgMLST v1.0 profile completion
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Query filters Select the fields for which you would like d interface. Solate Country Continent Solate Country Continent Solate Country Continent Solate Country Continent Solate Country C	oll up.	tery results. These will be available in the filters section of the query
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface. isolate country continent geage isovear_sampled isovear_sampled isovear_sampled isovear_asampled isovear_	oll up. tropdown lists containing known values on which to filter qu	ery results. These will be available in the filters section of the query private_project comments sender curator date_entered datestamp MLST profile completion Bexsero Antigen Sequence Typing (BAST) profile completion Cryptic Plasmid profile completion GC OMV peptide typing profile completion N. genorrhoeea cgMLST v1.0 profile completion N. meningitidis cgMLST v1.0 profile completion N. meningitidis cgMLST v1.0 profile completion N. GSTAP profile completion NG STAP profile completion
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scrue Query filters Select the fields for which you would like of interface. Select the field	oll up.	ery results. These will be available in the filters section of the query
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like of interface. Select the field	oll up.	ery results. These will be available in the filters section of the query private_project comments sender curator date_entered datestamp MLST profile completion Conjugative Plasmid profile completion Conjugative Plasmid profile completion Conjugative Plasmid profile completion Conjugative Plasmid profile completion N, gonorrheeae cgMLST v1.0 profile completion N, G MAST profile completion NG MAST profile completion CNG MAST Profile CNG Profile CN
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like of interface. isolate country continent continent geigion year date_sampled date_received date_received date_received con_culture epidemiological_year age_rrn age_rrn age_rrn age_mth sex disease cource epidemiology cource epidemiology cource epidemiology cource courc	oll up.	tery results. These will be available in the filters section of the query
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like of interface. Select the field	oll up.	ery results. These will be available in the filters section of the query private_project comments sender curator date_entered datestamp MLST profile completion Conjugative Plasmid profile completion Conjugative Plasmid profile completion Conjugative Plasmid profile completion Conjugative Plasmid profile completion N, gonorrheeae cgMLST v1.0 profile completion N, G MAST profile completion NG MAST profile completion CNG MAST Profile CNG Profile CN
Click the reset button to remove all user settin RESET ALL TO DEFAULTS Query filters' tab will scrue Query filters Select the fields for which you would like of interface. Select the fiel	oll up.	tery results. These will be available in the filters section of the query

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'.

ALL

NONE

DEFAULT

SET OPTIONS

11.13.6 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 in the 'Customise' section on the isolate contents page.

BIGSdb						
Home > Organisms > Neisseria spp. > Neisseria isolates						
Neisseria isolates database						
The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria spe database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequent sample.						
Query database	+) L	OG IN				
Search database	±. s					
Browse, search by any criteria, or enter list of attributes.	E P	ROJECTS	+			
	⊡ E	EXPORT +				
Search by combinations of loci This can include partial matches to find related isolates.	A	NALYSIS				
	¢ (USTOMISE	-			
	Locus Scher	ral options s display me display me field display				
	0 II	NFORMATION	+			
	S T	YPING				

Either select the locus id by querying for it directly.

BIGSdb		
Home > Organisms > Neisseria spp. > I	Neisseria isolates > Query locus information	
Query locus inform	ation	Help 🔀
Please enter your search criteria below	(or leave blank and submit to return all record	. Matching loci will be returned and you will then be able to update their display and query settings.
Search criteria		Display
id 🗸 =	 ✓ abcZ ✓ + 3 	Order by: id ascending v Display: 25 v records per page ()
Eilter query by	Action RESET SEARCH	
1 record returned. Click the hyperlink for	or detailed information.	
Locus options		
		late display* main display* query field* analysis* view
abcZ DNA partial CDS integer	433 🛛 🗹	ilele only 🗌 🗹 🗹

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.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Query locus information
Query locus information
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 Image: Constraint of the second
7 records returned. Click the hyperlinks for detailed informationCustomize Locus options
id data type locus type allele id format length length varies coding sequence isolate display" main display" query field" analysis" view
abcZ DNA partial CDS integer 433 Image: Constraint of the second
Once loci have been selected, click Customize 'locus options'.
scheme: MLST
7 records returned. Click the hyperlinks for detailed information. — Customize — Locus options

id	data type	locus type	allele id format	length	length varies	coding sequence	isolate display*	main display*	query field*	analysis*	view
abcZ	DNA	partial CDS	integer	433		\checkmark	allele only			\checkmark	
adk	DNA	partial CDS	integer	465			allele only			\checkmark	
aroE	DNA	partial CDS	integer	490		\checkmark	allele only			\checkmark	
fumC	DNA	partial CDS	integer	465		\checkmark	allele only		\checkmark	\checkmark	
gdh	DNA	partial CDS	integer	501		\checkmark	allele only			\checkmark	
pdhC	DNA	partial CDS	integer	480		\checkmark	allele only			\checkmark	
pgm	DNA	partial CDS	integer	450		\checkmark	allele only			\checkmark	
* Default values are displayed for this field. These may be overridden by user preference.											

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).

CHAPTER 12

User projects

If user projects are enabled, authenticated users are able to set up their own projects in order to group isolates for analysis. If private data and user quotas are enabled, these projects can include private records that can then be shared with other user accounts. Note, that simply adding a record to a user project does not make the record itself private.

Note: User projects can be enabled by an administrator by setting 'user_projects="yes" in the *config.xml* file for the database.

To create a new project, go to the isolate database contents page and expand the 'Projects' section. Click 'Your projects'.

BIGSdb							
Home > Organisms > Neisseria spp. > Neisseria isolates							
Neisseria isolates 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		DI LOG IN					
Search database							
Browse, search by any criteria, or enter list of attributes.		PROJECTS	-				
Search by combinations of loci		Public projects Your projects					
This can include partial matches to find related isolates.		EXPORT	+				
		MALYSIS					
		CUSTOMISE	+				
			+				

Enter the name of your project (this must be unique on the system - you will be told if the name is already used). You can optionally include a description - if you do this, this will appear within *isolate records* when accessed from your account. Click 'Create'.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects	
User projects	Letp
New private projects Projects allow you to group isolates so that you can analyse them easily together. Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an is record page (only you and users you give access will see this). New project Action Name: UK group W ST-11 2016	solate
Description: Group W ST-11 isolates from the UK in 2016. Existing projects	
You do not own or are a member of any projects.	

You can either add isolates to your project directly following a query or by manually editing a list of ids.

12.1 Adding isolates to a user project following a query

Perform an isolate database query. If you have set up a project, there will be a link in the results header box. Select your project and click 'Add these records'.

BIGSdb								
Home > Organisms > N	leisseria spp. > N	leisseria isolates > Sear	ch or browse database					
							💄 Help 📝 🚯 🚺] 🎤
Search or b	rowse da	atabase						
Enter search criteria or	leave blank to bi	rowse all records. Modify	form parameters to filter of	or enter a list of values				
Isolate provenar				Allele designation				
Combine with: AND				ST (MLST)		✓ 11		
country	v =	✓ UK	i + i			• 11	+ ()	
year	v =	✓ 2016		Display/sort optio	ns			
capsule group	✓ =	✓ W		Order by: id			✓ ascending ✓	
				Display: 25 🗸	records per page 👔			
RESET	EARCH	Click the hyperlinks for a	detailed information.					
	payea,	Boo	kmark query					
Your projects			kmark query					
				d bookmark				
—_Your projects UK group W ST-11	2016 V Add th			d bookmark				
—_Your projects UK group W ST-11	2016 V Add th	ese records 2020-0)7-29:1 Ad	d bookmark	MLST	Finetyping an	tigens	
Vour projects UK group W ST-11 (() () () 1 id isolate all	2016 ~ Add th 2 3 4 5 ases country	ese records 2020-0 6 () () isolate fields () year diseas)7-29:1 Ad) se specie	s capsule grou	p ST clonal complex		2 FetA_VR	
Vour projects UK group W ST-11 (() () () () () () () () () () () () ()	2016 ✓ Add th 2 3 4 5 ases country UK	ese records 2020-0 6 () () Isolate fields () year disea: 2016)7-29:1 Ad se specie Neisseria mer	es capsule grou ningitidis W	p ST clonal complex 11 ST-11 complex	PorA_VR1 PorA_VR 5 2	2 FetA_VR F1-1	
Your projects UK group W ST-11 (() () () () () () () () () () () () ()	2016 ✓ Add th 2 3 4 5 ases country UK UK [England	ese records 2020-0 6 (> (>) Isolate fields () 2016 2016 invasive (unspec) se specie Neisseria mer cified/other) Neisseria	<mark>s capsule grou</mark> hingitidis W hingitidis W	p ST clonal complex 11 ST-11 complex 11 ST-11 complex	PorA_VR1 PorA_VR 5 2 5 2	2 FetA_VR F1-1 F1-1	
Vour projects UK group W ST-11 (() () () () () () () () () () () () ()	2016 ✓ Add the 2 3 4 5 ases country UK [England UK [England	ese records 2020-0 6 (> (>) Isolate fields () year disea: 2016 d) 2016 invasive (unspec d) 2016 invasive (unspec d) 2016 invasive (unspec)7-29:1 Ad se specie Neisseria mer	is capsule grou ningitidis W ningitidis W	p ST clonal complex 11 ST-11 complex	PorA_VR1 PorA_VR 5 2	2 FetA_VR F1-1	

The records will be added to the project. Please note that it doesn't matter whether any of the records have been previously added.

12.2 Adding or removing isolates belonging to a user project by editing a list

From the user projects page, click the 'Add/remove records' link for the project that you wish to modify.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects
User projects
New private projects Projects allow you to group isolates so that you can analyse them easily together.
Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).
New project Action
Name: CREATE
Your projects
Delete Add/remove records Modify users Project Description Administrator Isolates Browse X Image: Comparison of the state of the
You can also add isolates to projects from the results of a query.
Note that deleting a project will not delete its member isolates.

Edit the list of ids of records belonging to the list. You can copy and paste this list if you wish to prepare it in a spreadsheet or text editor. Click 'Update' when finished.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects	
User projects	Help 🗹
Project: UK group W ST-11 2016 Group W ST-11 isolates from the UK in 2016. The list below contains id numbers for isolate records belonging to this project. You can add and remove records to this project by modifying the list of isolate ids. This only all records belong to the project - you will not remove isolate records from the database by removing them from this list. Isolate ids Action 42451 42475 42480 42480 42481 42485 42485 42485	ffects which

12.3 Accessing project isolates

To browse isolate records belonging to a project, go to the user projects page and click the 'Browse' link for the project.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects
User projects
New private projects
Projects allow you to group isolates so that you can analyse them easily together.
Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).
New project Action
Name: CREATE CREATE
Your projects
Delete Add/remove records Modify users Project Description Administrator Isolates Browse * * W group W ST-11 2016 Group W ST-11 isolates from the UK in 2016. Image: Comparison of the text of tex
You can also add isolates to projects from the results of a query.
Note that deleting a project will not delete its member isolates.

Alternatively, you can select the project from the *projects filter* on the isolate query page. This would enable you to combine the project with additional search criteria

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database	
Search or browse database	💄 Help 🗹 🟮 🕃 🗡
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.	
Isolate provenance fields Filters	
id v = v Enter value + () Publication	Select options
	i 1 selected
	Filter: Enter keywords
	Check all X Uncheck all
Sequence bir	
Include old record version	
Add filter: Display/sort options Action	✓ Add
	UKMenCar1
Order by: id vascending v	CUKMenCar2
Display: 25 v records per page (j	UKMenCar3

Please note that you will only see your project in the filter list if you are logged in. As a private project, only you will see it.

12.4 Allowing other users to share your project

You can share a project that you own with any other user. In order to do this, you must know the username that they use to log in with. They will see the project in their own list of private projects and in the query interface projects filter.

From the user projects page, click the 'Modify users' link for the specified project:

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects	
User projects	
New private projects Projects allow you to group isolates so that you can analyse them easily together.	
Projects allow you to group isolates so that you can analyse them easily together. Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).	
New project Action	
Name: CREATE	
Your projects	
Delete Add/remove records Modify users Project Description Administrator Isolates Browse X X VK group W ST-11 2016 Group W ST-11 isolates from the UK in 2016. Image: Comparison of the UK in 2016. Image: Comparison of the UK in 2016. Image: Comparison of the UK in 2016.	
You can also add isolates to projects from the results of a query.	
Note that deleting a project will not delete its member isolates.	

Enter the username of the person you wish to share with and click 'Add user':

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects	
User projects	la 🔽
Project: UK group W ST-11 2016	
User groups	
All members of selected user groups can view this project.	
Select user groups	
Select user groups able to access project	
Select options	
Update user groups	
Users	
The following users have permission to access the project (either explicitly or through membership of a user group).	
Users	
User Admin Add/Remove records Bloggs, Joe (jbloggs)	
Update users	
Add user	
Enter username: [hbratcher You need to know the username of any user you wish to add.	
Add user	

If user groups are in use, you can also share your project with all members of a user group from the same page.

Once a user has been added, you can give them permission to administer (add other users, modify the description or delete) or add/remove records to the project:

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects	
User projects	Help 🕜
Project: UK group W ST-11 2016	
User groups	
All members of selected user groups can view this project.	
Select user groups	
Select user groups able to access project	
Select options *	
Update user groups	
Users	
The following users have permission to access the project (either explicitly or through membership of a user group).	
Users	
Remove User Admin Add/Remove records Bloggs, Joe (jbloggs) Image: Comparison of the second se	
Update users	
Add user	
Enter username: hbratcher You need to know the username of any user you wish to add.	
tou need to know the username or any user you wish to add.	
Add user	

12.5 Deleting a user project

You can delete a project from the user projects page by clicking the 'Delete' link next to the project in question.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects	
User projects	
New private projects Projects allow you to group isolates so that you can analyse them easily together. Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on record page (only you and users you give access will see this). New project Action Name: CREATE	an isolate
Your projects Description Administrator Isolates Browse Image: State of the s	
You can also add isolates to projects from the results of a query. Note that deleting a project will not delete its member isolates.	

If the project contains any isolates you will be asked for confirmation. Click the 'Delete project' button.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > User projects
User projects
This project contains 196 isolates. Please confirm that you wish to remove the project (the isolates in the project will not be deleted). X Delete project
I New private projects
Projects allow you to group isolates so that you can analyse them easily together.
Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).
New project Action
Name: CREATE
Your projects
Delete Add/remove records Modify users Project Description Administrator Isolates Browse X X UK group W ST-11 2016 Group W ST-11 isolates from the UK in 2016. Isolates 196 Item to the tem tem tem tem tem tem tem tem tem te
You can also add isolates to projects from the results of a query.
Note that deleting a project will not delete its member isolates.

If the project contains no isolates, the confirmation page is skipped and the project is removed immediately.

Note: Removing a project does not delete the isolate records belonging to it. A project is simply a means of grouping records.

CHAPTER 13

Private records

Users with a status of 'submitter', 'curator', or 'admin' can upload private isolate records that will be hidden from public view. A quota needs to be set for the user by an admin before they are able to do this.

13.1 Uploading private records

In order to upload private records, you need to make sure that you are logged in. If your account has a quota, there will be a menu item called 'Private data'. Click the 'Upload/manage records' link.

BIGSdb			
Home > Organisms > Neisseria spp. >	Neisseria isolates		
<i>Neisseria</i> isolates	database		<u>ă</u>
	Query database	🕪 LOG OUT	+
Search database			
Browse, search by a	ny criteria, or enter list of attributes.	A PRIVATE DATA	
		PROJECTS	+
<mark>Search by combin</mark> This can include pa	nations of loci rrtial matches to find related isolates.	EXPORT	+
		🗠 ANALYSIS	
			+
			+

You will see an overview of your quota and links to upload and edit your records. Click the 'Upload private isolate records' link (assuming you have quota available).

BIGSd	b	
Home >	Organisms > Neisseria spp. > Neisseria isolates > Private records	
Priv	ate records	ď
Accourt	Limits Its have a quota for the number of private records that they can upload. Uploading of private data to some registered projects may not count against your quota. Records (total): 0 Records (quota): 0 Quota: 100 You can upload: 100	
1	Upload Upload private isolate records Curate Update private records <u>Curator's interface</u>	
	Note: This link takes you to the standard curator's interface. If you upload here, then your data will not be private. Either use the 'Upload' link above to upload private records to your quo (if you have any available), or use the upload link on the individual projects listed below to upload to those projects.	ota

You will then be taken to a curator page that allows you to upload data copy and pasted from a spreadsheet. There is a link to an Excel template that you will need to use to prepare your data - this is the same template used if you were to submit your data to a curator.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add isolate records	
Batch insert isolates	
Private data upload These isolates will count against your quota of private data.	
Quota available: 100	
 Instructions This page allows you to upload isolate data as tab-delimited text or copied from a spreadsheet. Pield header names must be included and fields can be in any order. Optional fields can be omitted if you wisk. Optionally enter references for your isolates as a semi-colon () separated list of PubMed ids (non-integer ids will You can also upload allele fields along with the other isolate data - simply create a new column with the locus of the will be added with a confirmed status and method set as 'manual'. You can choose whether or not to include an id number field - if it is omitted, the next available id will be used at Desplate Explates Explane Check the description of database fields for help with filling in the template. 	name (see the 'allowed_loci' tab in the Excel template for locus names).
Upload Paste in tab-delimited text (include a field header line).	Action

See batch adding isolate records for details of the upload process.

+

13.2 Modifying private records

Click the 'Upload/manage records' link on the contents page. BIGSdb Home > Organisms > Neisseria spp. > Neisseria isolates Neisseria isolates database 🕞 LOG OUT Query database **▲** SUBMISSIONS Search database Browse, search by any criteria, or enter list of attributes. A PRIVATE DATA **PROJECTS** Search by combinations of loci EXPORT This can include partial matches to find related isolates. MALYSIS **S** TYPING

Now click the 'Update private records' link. You will be taken to the curators' interface, where you will be able to use the standard curator tools.

Some > Organisms > Neisseria spp. > Neisseria isolates > Private records	
Cumits Accounts have a quota for the number of private records that they can upload. Uploading of private data to some registered projects may not count against your quota. Records (total): Records (quota): Quota: 100 You can upload: 100 Upload U	
Accounts have a quota for the number of private records that they can upload. Uploading of private data to some registered projects may not count against your quota. Records (quota): Quota: 100 You can upload: 100 Upload Upload Upload Lupload private isolate records	Help 🔽
Records (total): 0 Records (quota): 0 Quota: 100 You can upload: 100 Upload Upload private isolate records	
You can upload: 100 Upload Upload private isolate records	
Upload private isolate records	
Curate	
Update private records Curator's interface	
Note: This link takes you to the standard curator's interface. If you upload here, then your data will not be private. Either use the 'Upload' link above to upload private records to you (if you have any available), or use the upload link on the individual projects listed below to upload to those projects.	ir quota

Use the curators' interface to make any changes to your isolate records, including uploading genome data.

13.3 Sharing access to private records

If *user projects* are enabled on the database, you can share access to your private isolates by adding them in to a user project and then sharing this.

See user projects for more details.

CHAPTER 14

Data analysis plugins

14.1 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 'Analysis' section on the main contents page.

BIGSdb			
Home > Organi	ms > Neisseria spp. > Neisseria isolates		
Neisser	<i>ia</i> isolates database		
	Query database	→3 LOG IN	
	Search database		
	Browse, search by any criteria, or enter list of attributes.	PROJECTS	+
		EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.	MALYSIS	
		CUSTOMISE	+
			+

Jump to the 'Analysis' category, follow the link to BLAST, then click 'Launch BLAST'.

BLAST

Summary: BLAST a query sequence against selected isolate data

The BLAST plugin enables you to BLAST a sequence against any of the genomes in the database, displaying a table of matches which can be downloaded in Excel format. In addition, the matched sequence regions are made available for download in FASTA format, either with or without surrounding flanking sequence.



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.

13	139M	B99; NIBSC 2795; Z1099	Philippines	1968	intusite (unspecifico) othery	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
	0120/95	b)), ((b)C_2/)), 210/)			invasive (unspecified/other)		X	117	51-1 comptex	5-2	14	15-
5	1		Germany	1999	carrier	Neisseria meningitidis	E	864				
6	2		Germany	1999	carrier	Neisseria meningitidis	В	854	ST-18 complex			
7	3		Germany	1999	carrier	Neisseria meningitidis	w		ST-174 complex			
8	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
9	S3131	B213; NIBSC 2813; Z1213	Ghana		invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-3
0	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			
		B227; NIBSC 2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex	5-1	9	F3-1
24	S4355	DZZ/, NIDSC_2000, Z122/										
	\$4355 9	B227, NIBSC_2000, 21227	Germany	1999	carrier	Neisseria meningitidis	В	930				
	9	1 2 3 4 5 6 tools	Germany	1999	carrier		В					
es (« Ar	9 2) (•)	1 2 3 4 5 6 tools		1999			В					
es @	9	1 2 3 4 5 6 tools	Germany	1999	carrier	Neisseria meningitidis	В					

Select the isolate records to analyse (on large databases you will need to enter a list of ids). 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.

ione - organisins - neissena spp ne	leisseria isolates > Plugins > BLAST	
BLAST	Help	
Please select the required isolate ids to B	BLAST against (use CTRL or SHIFT to make multiple selections) and paste in your query sequence. Nucleotide or peptide sequences can be queried.	
Isolates Isolates	Paste sequence	
	AAGCCTTAGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCGGCAAGCCTGAA	
	AAAATCGCCAACGACATCCGCTGGCAGCGGCCGGCGCGCGC	
	GCTGCCAAGTGTTCGGCAACGACGTTACCATCGGTATGGCGGGGGCGCGTCGGGCAATTTCGAGCTGAACGT TATATGCCCGTTATCGCCTACAACCTCTTGCAATCCATCC	
6 - AC	CGAACACTGCGCCATCGGCATCGAACCCGTACCGGAAAAAATCGACTATTTCCTGCACCATTCCCTGATG	
	TCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAC	
Clear List all		
Include in results table	Parameters	
country	▲ BLASTN word size: 11 × ()	
region year	BLASTN scoring: reward:2; penalty:-3; gap open:5; gap extend:2 👻 👔	
date sampled	Hits per isolate: 1	
isoyear sampled week sampled	Flanking length (bp): 100 v ()	
date received		
non culture epidemiological year	Options Restrict included sequences by	
age yr	Show isolates with no matches Sequence method: Constant position in FASTA Droiect:	
	i i i i i i i i i i i i i i i i i i i)
	Experiment: 🗸 🗸	
Action		

Click submit. If you are querying against 10 or fewer genomes then the results are run immediately, otherwise the job is sent to the job queue.

A table of BLAST results will be displayed.

> Or	ganisms > I	veisseria spp.	> Neisseria iso	lates > Job	status v	/iewer						
o st	atus v	viewer										
	_											
	Status											
			db_009799_159		173							
		t time: 2020 Status: finish	-07-29 15:33:30)								
			-07-29 15:36:2:	L								
		ogress: 100										
		time: 2020 Ltime: 7 sec	-07-29 15:36:29	9								
	IOCA	tume: 7 sec	onas									
itpu	t											
•		9/ identity A		Viewerteler		Carlor	Ctort	Fed	Onionstation	E un luc	Dit seens	
late id	i Isolate A4/M1027		lignment lengtl 465	n Mismatche 8	s Gaps 0	180177	Start 19444	End 19908 extract →	Orientation →	0.0	803	
2	120M	98.280	465	8	0	180583	4782	5246 extract →	←	0.0	803	
7	7891	98.280	465	8	0	180965	19869	20333 extract →	\rightarrow	0.0	803	
10	6748 129E	98.280 98.280	465 465	8	0	181406 181867	36669 35889	37133 extract →	\rightarrow	0.0	803 803	
11 13	129E	98.280	465	8	0	181867	36775	36353 extract → 37239 extract →	\rightarrow \rightarrow	0.0	803	
19	53131	98.280	465	8	0	182318	19090	19554 extract →	\rightarrow	0.0	803	
24	S4355	98.280	465	8	0	215673	4534	4998 extract →	-	0.0	803	
30	14	99.355	465	3	0	8	1363524	1363988 extract →	\rightarrow	0.0	825	
31	10	98.280	465	8	0	182380	5559	6023 extract →	\rightarrow	0.0	803	
34	20	98.280	465	8	0	182815	19783	20247 extract →	\rightarrow	0.0	803	
35	26	98.280	465	8	0	182880	18879	19343 extract →	\rightarrow	0.0	803	
52 61	243 393	98.280 98.280	465 465	8	0	183381 183648	4532 7217	4996 extract → 7681 extract →	\rightarrow	0.0	803	
64	254	98.280	465	8	0	183818	4716	5180 extract →	 	0.0	803	
67	S5611	98.280	465	8	0	184143	4804	5268 extract →	÷	0.0	803	
82	11-004	98.280	465	8	0	184297	19518	19982 extract →	\rightarrow	0.0	803	
84	IAL2229	98.280	465	8	0	184635	4530	4994 extract →	←	0.0	803	
90	CN100	98.280	465	8	0	184814	5350	5814 extract →	<i>←</i>	0.0	803	
es												
Z	FASTA			X	FAST	A with flar	nking		Tabl	e (tab-de	imited text)	
x	Table (Ex	cel format)				ile containi	ng all outp	out				
					files							

Clicking any of the 'extract' buttons to display the matched sequence.

Output											
Isolate id	Isolate	% identity	Alignment length	Mismatches	Gaps	Seqbin id	Start	End	Orientation	E-value	Bit score
1	A4/M1027	98.280	465	8	0	180177	19444	19908 extract →	\rightarrow	0.0	803
2	120M	98.280	465	8	0	180583	4782	5246 extract →	←	0.0	803
7	7891	98.280	465	8	0	180965	19869	20333 extract →	\rightarrow	0.0	803
10	6748	98.280	465	8	0	181406	36669	37133 extract →	\rightarrow	0.0	803
11	129E	98.280	465	8	0	181867	35889	36353 extract →	\rightarrow	0.0	803
13	139M	98.280	465	8	0	182004	36775	37239 extract →	\rightarrow	0.0	803
19	S3131	98.280	465	8	0	182318	19090	19554 extract →	\rightarrow	0.0	803
24	S4355	98.280	465	8	0	215673	4534	4998 extract →	←	0.0	803
30	14	99.355	465	3	0	8	1363524	1363988 extract →	\rightarrow	0.0	825
31	10	98.280	465	8	0	182380	5559	6023 extract →	\rightarrow	0.0	803
34	20	98.280	465	8	0	182815	19783	20247 extract →	\rightarrow	0.0	803

The extracted sequence is shown along with a translated sequence and flanking sequences.

	BIGSdb
I	Home > Organisms > Neisseria spp. > Neisseria isolates > Extracted sequence
	Extracted sequence: Seqbin id#:180965 (19869-20333)
	Sequence position Options sequence bin id: 180965 Flanking sequence length: 100 マ sequence method: Solexa Reload start: 19869 19869 end: 20333 Iength: 465 orientation: →
	Sequence
	R H G F E Q P S R I R R K S R R Q T R R I V R L A V C Q R A E Q I G T G L N S H P E Y A E K A A A K L A E L S G L P F V S A P N K F F2 A R V T A I P N T P K K P P P N S P N C P A C R L S A R R T N L F3 1 CGGCACGGGTTTGAACAGCCGAATAGCGCGCGAAACTGCCGCAATTGTCCGGCTGCTGCTGTCGGCGCGGACAAATT 100

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.

84 IAL2229 98.280 465 8 0 184635 4530 4994 extract - ← 0.0 803 90 CN100 98.280 465 8 0 184814 5350 5814 extract - ← 0.0 803	67 82	S5611 11-004	98.280 98.280	465 465	8	0	184143 184297	4804 19518	5268 extract → 19982 extract →	$\stackrel{\leftarrow}{\rightarrow}$	0.0	803 803	
Files FASTA FASTA FASTA with flanking Table (tab-delimited text) Table (Excel format) Tar file containing all output						-		4530	4994 extract →				
			cel format)			Tar f	ile containi	-		Tabl	e (tab-del	imited tex	t)
	Please not	e that job re	esults will re	main on the serv	er for 7 days.								

14.1.1 Include in results table fieldset

This selection box allows you to choose which isolate provenance fields will be included in the results table.

country	
region	
year	
date sampled	
isoyear sampled	
week sampled	
date received	
non culture	
epidemiological year	
age yr	

Multiple values can be selected by clicking while holding down Ctrl.

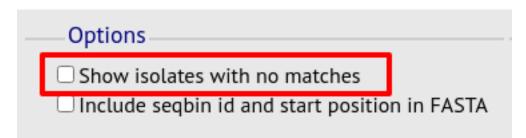
14.1.2 Parameters fieldset

This section allows you to modify BLAST parameters. This affects sensitivity and speed.

Parameters	
BLASTN word size:	11× (j
BLASTN scoring:	reward:2; penalty:-3; gap open:5; gap extend:2 🖌 👔
Hits per isolate:	1 🗸
Flanking length (bp):	100 ~ (i)
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.

14.1.3 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.

14.1.4 Filter fieldset

This section allows you to further filter your collection of isolates and the contig sequences to include.

Restrict included seq	uences by		
Sequence method:		* (i)	
Project:			× (i)
Experiment:	~ (i)		U

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.

14.2 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.

Third party: GrapeTree iTOL PhyloViz Microreact

12	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningitidis	s B	3 1015	ST-32 complex	7	16	
13	139M	B99; NIBSC_2795; Z1099	Philippines	1968		Neisseria meningitidis	s A	A 1	ST-1 complex	5-2	10	F5-1
14	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitidi:	s X	K 117			14	
15	1		Germany	1999	carrier	Neisseria meningitidi:	s E	864				
16	2		Germany	1999	carrier	Neisseria meningitidi	s B	854	ST-18 complex			
17	3			1999	carrier	Neisseria meningitidis			ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningitidis	s B	3 19	ST-18 complex			
19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	s A	4 4	ST-4 complex	7	13-1	F1-5
20	5		Germany	1999	carrier	Neisseria meningitidis			ST-198 complex			
21	6		Germany	1999	carrier	Neisseria meningitidis	s N	G 198	ST-198 complex			
22	7		Germany	1999	carrier	Neisseria meningitidis	s E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningitidis	s B	3 32	ST-32 complex			
24	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidi	s A	A 5	ST-5 complex	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria meningitidi	s B	3 930	ST-334 complex			
			,									
«) 🕢	1 2 3 4 5 6 tools	00)								
(« Ar) 🔇 nalysis 1 💺 Breakdo	tools	Combinations	Poly	morphic sites Publication	Sequence bin						
(« Ar	Breakdo	tools	Combinations Gene Presence		morphic sites Publication me Comparator BLAST		CR					

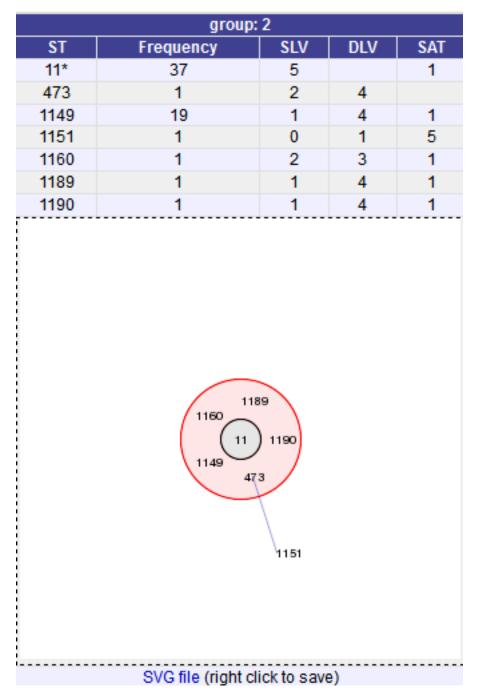
If there multiple schemes that can be analysed, these can then be selected along with the group definition.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > BURST	
BURST analysis	Help 🗹
This is the original BURST algorithm, developed by Ed Feil, first implemented by Man-Suen Chan. This version has bee Jolley.	en adapted for use as a plugin for the BIGSdb database software by Keith
BURST analysis can be used to:	
 Divide strains into groups according to their allelic profiles. Count the number of Single Locus Variants (SLV), Double Locus Variants (DLV) and Satellites (SAT) for each sequ Identify the potential Ancestral Type (AT). These are shown with an asterisk next to their names in the results ta 	
Graphic representations of BURST groups can be saved in SVG format. This is a vector image format that can be mani Inkscape.	ipulated and scaled in drawing packages, including the freely available
Options	Action
Select scheme: MLST	SUBMIT
Group definition: profiles match at $n-2$ v loci to any other member of the group (n = number of loci in scheme).	
☑ Shade variant rings □ Hide variant names (useful for overview if names start to overlap)	

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.

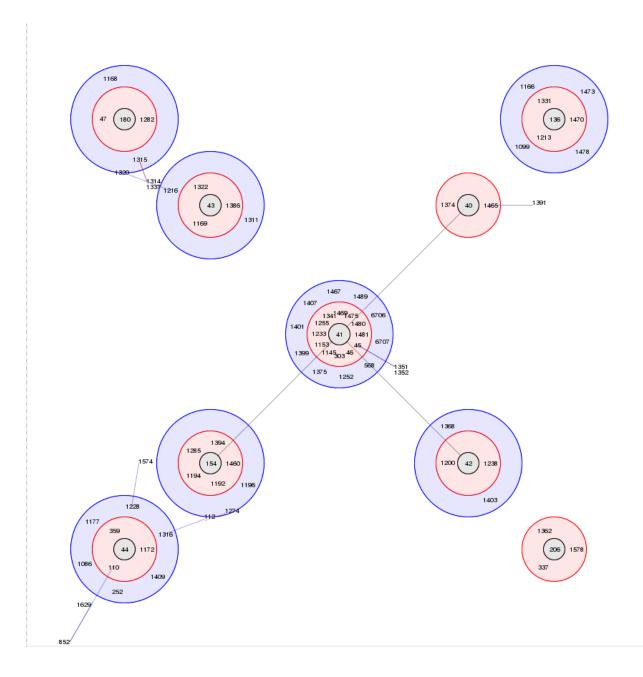
A series of tables will be displayed indicating the groups of profiles. Where one profile can be identified as a central genotype, i.e. the profile that has the greatest number of other profiles that are single locus variants (SLV), double locus variants (DLV) and so on, a graphical representation will 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
		1100 114	8	
	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.



14.3 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 accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb

Home > Organisms > Neisseria spp. > Neisseria isolates Neisseria isolates database

Query database	D LOG IN	
Search database		
Browse, search by any criteria, or enter list of attributes.		+
	EXPORT	+
Search by combinations of loci This can include partial matches to find related isolates.	ANALYSIS	
nns can nicuue partiat marches to nnu related isolates.		+
		+

Jump to the 'Analysis' category, follow the link to 'Codon Usage', then click 'Launch Codon Usage'.

The function can be selected by clicking the 'Codon usage' link in the Analysis section of the main contents page.

Analysis
1 1141 9 13
Analysis plugins - Jump to: Codon Usage Gene Presence Genome Comparator BLAST rMLST species identity PCR
Codon Usage
Summary: Determine codon usage for specified loci for an isolate database query
The codon usage plugin calculates the absolute and relative synonymous codon usage by isolate and by locus for any dataset or the whole database. Specific loci or the loci that are members of a particular scheme can be chosen for analysis.
Documentation bigsdb readthedocs.io 🗹

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.

1 1			Germany	1999	carrier	r	Neisseria meningit	tidis	E	864				
L 6 2			Germany	1999	carrier	r	Neisseria meningit	tidis	В	854	ST-18 complex			
7 3			Germany	1999	carrier	r	Neisseria meningit	tidis	W	174	ST-174 complex			
8 4			Germany	1999	carrier	r	Neisseria meningit	tidis	В	19	ST-18 complex			
9 S3131	B213; NIBS	C_2813; Z1213	Ghana	1973 invasiv	ve (unspeci	ified/other)	Neisseria meningit	tidis	A	4	ST-4 complex	7	13-1	F1-5
20 5			Germany	1999	carrier	r	Neisseria meningit	tidis	NG	198	ST-198 complex			
21 6			Germany	1999	carrier	r	Neisseria meningit	tidis	NG	198	ST-198 complex			
2 7			Germany	1999	carrier	r	Neisseria meningit	tidis	E	60	ST-60 complex			
23 8			Germany	1999	carrier	r	Neisseria meningit	tidis	В	32	ST-32 complex			
4 S4355	B227; NIBS	C_2806; Z1227	Denmark	1974 invasiv	ve (unspeci		Neisseria meningit		A	5	ST-5 complex	5-1	9	F3-1
									В	930	ST-334 complex			
25 9			Germany	1999	carrier	r	Neisseria meningit	tials	D	930	SI-354 complex			
(() () Analysis		34	Sermany	1999	carrier	r	Neisseria meningit	tidis	D	930	31-334 complex			
© ()		3 4	Combinations	Polymorphi		Publications	_	tidis	D	950	S1-334 Complex			
(() () Analysis () Break			00		iic sites F	Publications	_	PCR	D	026	S1-334 Complex			

Third party: GrapeTree iTOL PhyloViz Microreact

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.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Codon usage	
Codon usage analysis	
included. It is important to note that correct identification of codons can only be achieved from the sequence bin will not be analysed. Please check the loci that you would like to in	
Select idsLoci	Schemes
Paste in list of ids to include, start a new line for each. Leave blank to include all lds. 1 1 2 3 4 5 All None Paste list	Image: Second
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: alphabetical C or G ending codons first
Do not include sequences with problem flagged (defined alleles will still be used)	

Click submit. The job will be submitted to the queue and will start running shortly.

Output files will be created in both tab-delimited text and Excel formats for the following:

- 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

BIGSdb			
Home > Organisms > Neisseria spp. > Neisseria isolates > Job s	tatus viewer		
Job status viewer			
Status Job Id: BIGSdb_004053_1596101846_944 Submit time: 2020-07-30 09:37:26 Status: finished Start time: 2020-07-30 09:37:35 Progress: 100% Stop time: 2020-07-30 09:38:00 Total time: 24 seconds	417		
Output Files			
Absolute frequency of codon usage by isolate (text)	Absolute frequency of codon usage by locus (text)	Relative synonymous codon usage (RSCU) by isolate (text)	
Relative synonymous codon usage (RSCU) by locus (text)	Absolute frequency of codon usage by isolate (Excel)	Absolute frequency of codon usage by locus (Excel)	
Relative synonymous codon usage (RSCU) by isolate (Excel)	Relative synonymous codon usage (RSCU) by locus (Excel)	Tar file containing all output files	
(RSCU) by locus (text)	by isolate (Excel)	usage by locus (Excel)	

14.4 Field breakdown

The field breakdown plugin for isolate databases displays the frequency of each value for fields, alleles and schemes. The function can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb				
Home > Organis	ns > Neisseria spp. > Neisseria isolates			
Neisser	<i>ia</i> isolates database			
	Query database		DI LOG IN	
	Search database			
	Browse, search by any criteria, or enter list of attributes.		PROJECTS	+
			EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.		🗠 ANALYSIS	
	This can include partial matches to initi related isolates.			+
				+

In the 'Breakdown' category, follow the link to 'Field Breakdown', then click 'Launch Field Breakdown'.

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.



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.

17												
	3			1999	carrier	Neisseria meningitidis		174				
18	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973 invasiv	e (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	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			
24	S4355	B227; NIBSC_2806; Z1227	Denmark	1974 invasiv	e (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria meningitidis	В	930	ST-334 complex			
			\sim									
Aı	nalysis	tools	00									
	5		Combinations	Publication	s Sequence bin							
	5	own: Fields Two Field		Publication Comparator		ecies id PCR						
	Breakd	own: Fields Two Field ysis: Codons Gene Preser				ecies id PCR						

A chart will be displayed for the first field. Depending on the field type, this may be either a world map (for country or continent fields), pie chart, or bar chart.

BIGSdb		
Home > Organisms > Neisseria spp. > N	Veisseria isolates > Plugins > Field breakdown	
Field breakdown of	dataset	Help 🗹
Isolate records: 66,602 Field selection Select field: Country List: © fields O loci O schemes		Controls
	5.000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.0000 3.00000 3.00000 3.00000000	

Other fields can be chosen by selecting them in the dropdown list box.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Field breakdown	
Field breakdown of dataset	Help 🗹
Islate records: 66,62 Select field country List: Pields Loci Schemes	Controls Theme: Range: Projection: Natural Earth

You can also breakdown loci and schemes by clicking the appropriate button. This will re-populate the dropdown list.

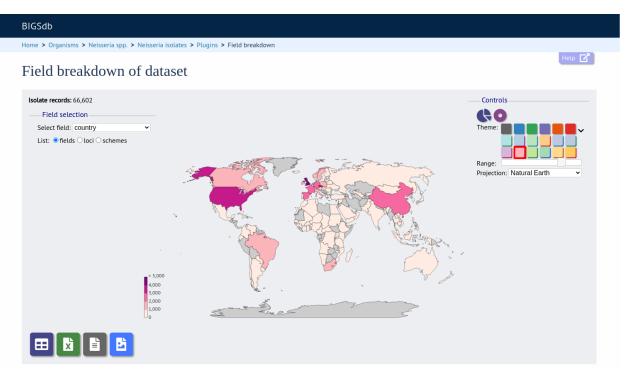
BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Field breakdown	
Field breakdown of dataset	Help 🗹
Isolate records: 66,602 Field selection Select field: clonal_complex (MLST) ✓ List: ○fields ○ loci @ schemes	Max segments: 20
clonal_complex (55 values)	
11.2% 26.6% 12.5% 12.4%	
No value ST-11 complex ST-41/44 complex ST-32 complex ST-269 complex ST-269 complex ST-213 complex ST-22 complex ST-269 complex ST-103 comple	

The charts are dynamic and you can manipulate some aspects of them using controls shown on the screen.

14.4.1 Maps

World maps are shown for country and continent fields (provided standardized country names are used in the database). The maps can be modified in a number of ways.

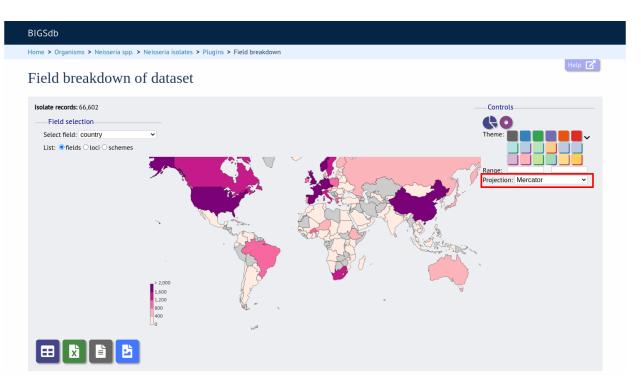
The colour theme can be changed by clicked the appropriate coloured square. Additional themes are available by clicking the '>' link.



The range that is used to decide the colour boundaries can be changed by using the range slider.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Field breakdown	
Field breakdown of dataset	Help 🔽
Field selection Select field; country List: ● fields □ loci □ schemes	Controls Theme: Range: Projection: Natural Earth

Finally the map projection can be changed. The default 'Natural Earth' provides a reasonable display for most latitudes but you may prefer others such as 'Mercator'.

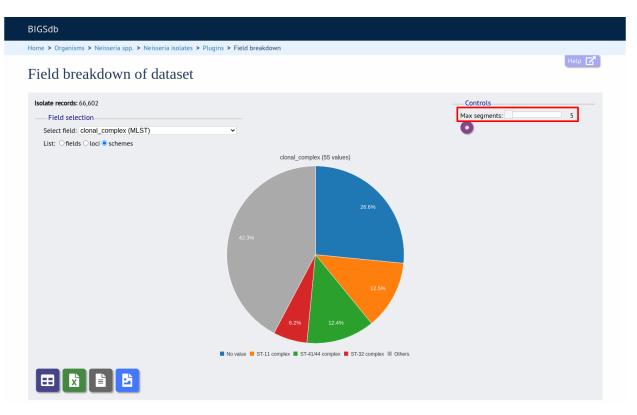


To see the same data as a pie chart, click the 'Pie' or 'Donut' icons.



14.4.2 Pie charts

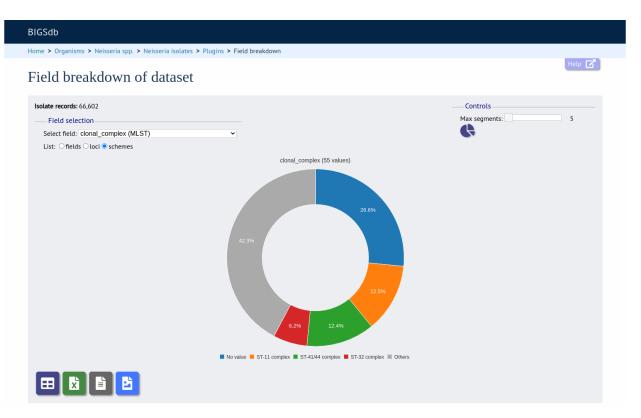
The maximum number of segments shown can be modified by sliding the 'Max segments' control. Low frequency values will be grouped in to a segment called 'Others'.



The chart can be transformed in to a donut chart by clicking the donut icon.

Controls	
Max segments:	5

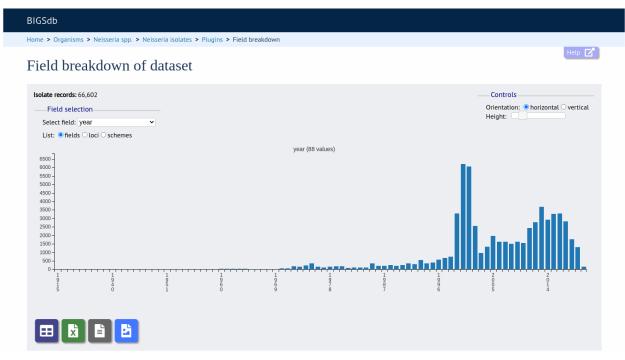
The icon changes to a pie chart image (clicking this will return to the pie chart).



Values can also be removed from the analysis by clicking their label in the legend below the chart. The percentages of the other values will be recalculated. Clicking the label again will re-add the value.

14.4.3 Bar charts

Integer fields will be displayed as a bar chart.



You can modify the height and the orientation of the chart using the controls.

14.4.4 Line charts

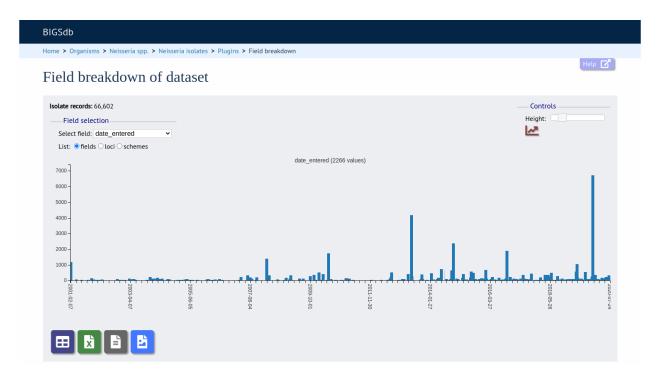
Date fields will be displayed as a line chart. By default this shows the cumulative values.

BI	GSdb						
Hor	ne > Organisms > Neisseria spp. > Neisseri	ria isolates > Plugins > Field breakdov	wn				
F	ield breakdown of da	itaset					Help 🔽
1:	solate records: 66,602 Field selection Select field: date_entered v List: @ fields Oloci Oschemes					Controls Height:	
	7		date_entered (2266 value	es)			
	70000 - 60000 - 50000 - 40000 - 30000 - 20000 - 10000 -						
		2005-06-05			2014-01-27	2018-05-26	2020-07-24
1	60000 - 50000 - 40000 - 20000 - 10000 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0	2007-08-04				-2018-05-26 	

The chart can be converted in to a bar chart showing discrete values by clicking the bar chart icon.



The icon changes to a line chart image (clicking this will return to the line chart).



14.4.5 Summary tables

The field breakdown can be displayed as a summary table containing values and percentages of all values. This can be selected by clicking the table icon below the displayed chart.



The table can be re-ordered by clicking any of the headings.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Field breakdown	
Breakdown by country	Help 🔽

country	¢	Frequency +	Percentage +
UK [England]		12836	19.27
UK [Scotland]		5228	7.85
UK		3730	5.60
USA		3263	4.90
Czech Republic		3112	4.67
France		2983	4.48
Germany		2769	4.16
Spain		2273	3.41
China		2185	3.28
Norway		1819	2.73
The Netherlands		1752	2.63
UK [Wales]		1670	2.51
Sweden		1540	2.31
Canada		1477	2.22
South Africa		1321	1.98
Italy		1285	1.93
Ireland		1193	1.79

The same table can be exported as an Excel file by clicking the Excel icon.



Alternatively, it can be exported as a tab-delimited text file by clicking the text file icon.

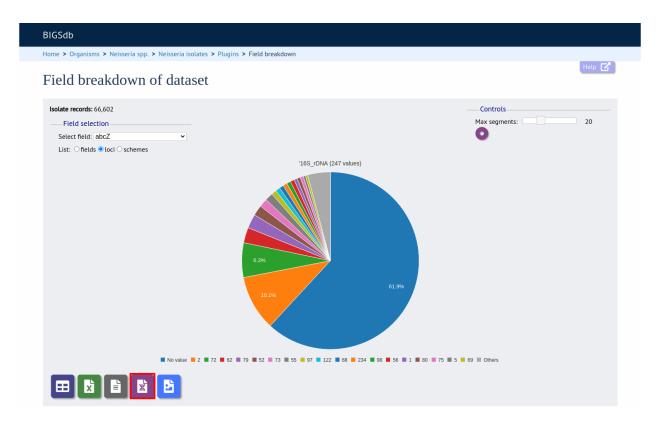


The chart image can also be saved as a SVG file, suitable for manipulation in a graphics program.



14.4.6 Exporting allele sequences

If a locus breakdown is being display, you can choose to export the allele sequences in FASTA format by clicking the FASTA file icon.



14.5 Gene Presence

The Gene Presence analysis tool will determine whether loci are present or absent, incomplete, have alleles designated, or sequence regions tagged for selected isolates and loci. If a genome is present and a locus designation not set in the database, then the presence and completion status are determined by scanning the genomes. The results can be displayed as interactive pivot tables or a heatmap. The analysis is limited to 500,000 data points (locus x isolates).

The function can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb Home > Organisms > Neisseria spp. > Neisseria isolates Neisseria isolates database DI LOG IN Query database **1** SUBMISSIONS Search database Browse, search by any criteria, or enter list of attributes. **PROJECTS** Z EXPORT + Search by combinations of loci MALYSIS This can include partial matches to find related isolates. CUSTOMISE + **1** INFORMATION **S** TYPING

Jump to the 'Analysis' category, follow the link to 'Gene Presence', then click 'Launch Gene Presence'.

Gene Presence			
Summary: Analyse presence/absence of loci for dataset generated from query results			
The Gene Presence analysis tool will determine whether loci are present or absent, incomplete, have alleles designated, or sequence regions tagged for selected isolates and loci. If a genome is present and a locus designation not set in the database, then the presence and completion status are determined by scanning the genomes. The results can be displayed as interactive pivot tables or a heatmap. The analysis is limited to 500,000 data points (locus x isolates).			
Documentation bigsdb.readthedocs.io 🗹			
BIGSdb Insec > Openties > Netoria gg, > Netoria Islans > Pagin > Gee pressue			
Gene Presence - Neisseria isolates			
Heatmap	Owng malyin Cartols Cartols Farmer		
etter (C210) Etter (C210) Etter (C210)			
Launch 'Gene Presence'			

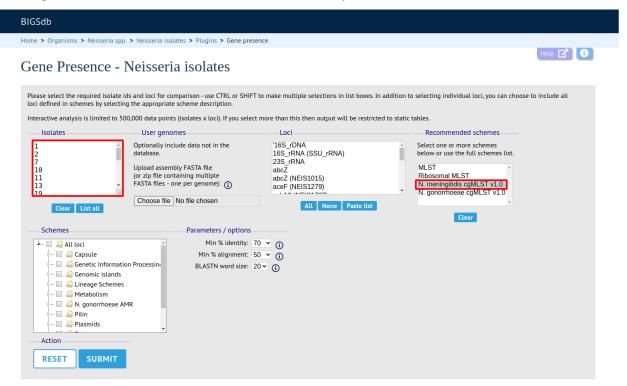
Alternatively, it can be accessed following a query by clicking the 'Gene Presence' button at the bottom of the results table. Isolates returned from the query will be automatically selected within the plugin interface.

61	393	B392; NIBSC_2823; Z1392	Greece	1968	carrier	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
64	254	B439; NIBSC_2812; Z1439	Djibouti	1966	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F1-7
67	S5611	B466; NIBSC_2765; Z1466	Australia	1977	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
82	11-004	B503; NIBSC_2826; Z1503	China	1984	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-8
84	IAL2229	B506; NIBSC_2816; Z1506	Brazil	1976		Neisseria meningitidis	A	5	ST-5 complex	20	9	F2-1
90	CN100	B534; NIBSC_2768; Z1534	UK [England]	1941	invasive (unspecified/other)	Neisseria meningitidis	Α	21		5-2	10	F3-9
120	F4698	NIBSC_2731; Z3515	Saudi Arabia	1987	carrier	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1
128	F6124	NIBSC_2730; Z3524	Chad	1988	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1
160	1014	NIBSC_2821; Z3667	Sudan	1985	invasive (unspecified/other)	Neisseria meningitidis	Α	4	ST-4 complex	7	13-1	F1-5
210	H1964	NIBSC_2796; Z3771	UK	1987	invasive (unspecified/other)	Neisseria meningitidis	А	5	ST-5 complex	20	9	F3-1
237	H44/76	44/76-3; NIBSC_2724; Z3842	Norway	1976	invasive (unspecified/other)	Neisseria meningitidis	В	32	ST-32 complex	7	16	F3-3
238	153	NIBSC_2733; Z3905	China	1966	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1
@	0	1 2 3 4 🕥	>>									

Analysis tools										
🕓 Breakdown:	Fields	Two Field	Combinatio	ons	Polym	orphic sites	Publicatio	ns	Sequence bin	
🛃 Analysis:	BURST	Codons	Gene Preser	nce	Genom	e Comparato	r BLAST	rt	MLST species id	PCR
🐻 Export:	Dataset	Contigs	Sequences							
🗹 Third party:	GrapeTr	ee iTOL	PhyloViz	Micro	oreact					

Select the isolates to include. Analysis can be performed on any selection of loci, or more conveniently, you can select a scheme in the scheme selector to include all loci belonging to that scheme. You can also select a recommended scheme if these have been defined.

The parameters of the BLAST query used to determine presence or absence can be modified, but in most cases the default options should work well. Click 'Submit' to start the analysis.

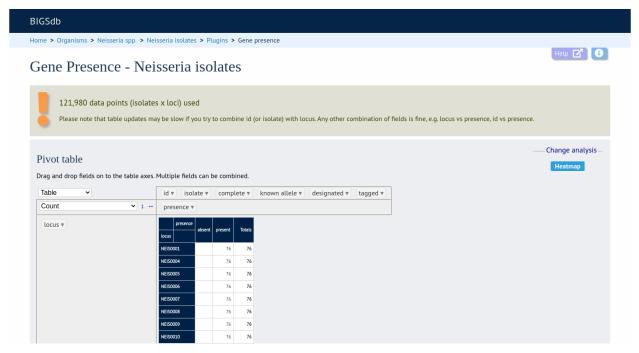


The job will be sent to the job queue. When it has finished, you will have two options to display the output: 'Pivot Table' or 'Heatmap'.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Job status viewer	
Job status viewer	
Status Job id: BIGSdb_017561_1596120664_37507 Submit time: 2020-07-30 14:51:04 Status: finished Start time: 2020-07-30 14:51:13 Progress: 100% Stop time: 2020-07-30 14:52:46 Total time: 1 minute and 32 seconds	
Output Pivot Table Heatmap Files	
Presence/absence (Excel) Presence/absence (text) Tar file containing all output files	
Please note that job results will remain on the server for 7 days.	

14.5.1 Pivot Table

Clicking the 'Pivot Table' button will display an interactive pivot table. The default display shows the number of isolates for which each locus is present or absent.



You can break down any combination of fields by dragging them from the field area at the top of the table to either of the axes. For example, to show how many isolates have alleles designated and sequence regions tagged for each locus, drag the 'designated' and 'tagged' fields to the x-axis selector.

BIGSdb		
łome > Organisms > Neisseria spp. > Neis	seria isolates > Plugins > Gene presence	
Gene Presence - Neis	sseria isolates	Help 🗹 🚺
121,980 data points (isolates Please note that table updates may	x loci) used / be slow if you try to combine id (or isolate) with locus. Any other combination of fields is fine, e.g. locus vs presence, id vs presence.	
Pivot table Drag and drop fields on to the table axes. M Table	Aultiple fields can be combined. id v isolate v complete v known allele v designated v tagged v	— Change analysis — Heatmap
	presence usern present Totals NEISCOOL C 76 76 NEISCOOL C 76 76	

The table will be re-drawn including these fields.

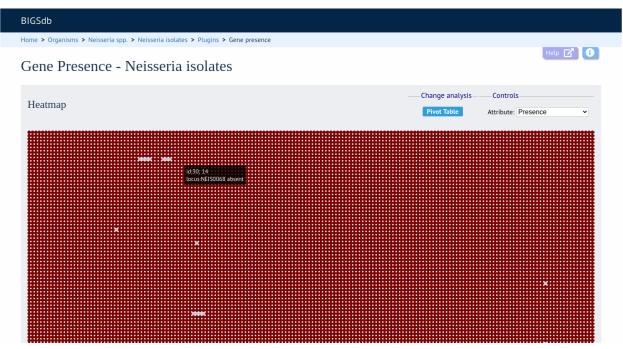
lGSdb		
ome > Organisms > Neisseria spp. > N	eisseria isolates > Plugins > Gene presence	
Gene Presence - Ne	isseria isolates	Help 📝 🚺
121,980 data points (isolat Please note that table updates r	es x loci) used hay be slow if you try to combine id (or isolate) with locus. Any other combination of fields is fine, e.g. locus vs presence, id vs	presence.
Pivot table Drag and drop fields on to the table axe Table	id v isolate v complete v known allele v	—— Change analysis— Heatmap
locus v	presence absent present	
	designated not designated not designated Totals	
	tagged untagged untagged tagged untagged tagged untagged	
	NEIS0001 76 76	
	NEIS0004 67 9 76	
	NEISO005 71 5 76	
	NEISO06 76 76	
	NEISO007 76 76	
	NEIS0008 76 76	

Note: If your dataset has more than 100,000 data points (locus x isolates), then be aware that combining both id (or isolate) and locus within the table will result in sluggish performace. Any other combination of fields should be fine.

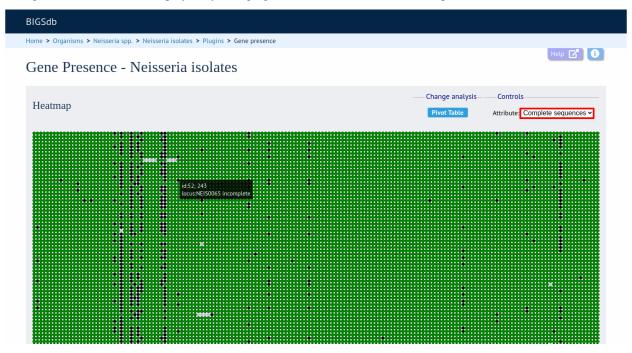
14.5.2 Heatmap

Clicking the 'Heatmap' button will display an interactive heatmap. By default the display shows the presence or absence of a locus for each isolate.

Hovering the mouse cursor or touching a region will identify the isolate and locus in a tooltip.



Change the attribute that is displayed by changing the selection in the attribute dropdown box:



The heatmap does scale to the number of records required to be displayed. If you find individual points to be too small, then choose a smaller subset of data to display:

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria is	olates > Plugins > Gene presence	
Gene Presence - Neisser	ia isolates	Help 🔽 🚺
Heatmap		Change analysis - Controls Pivot Table Attribute: Complete sequences
	fi224: 54355 Tracus VEI S0044 complete	

14.6 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 by selecting the 'Analysis' section on the main contents page.

BIGSdb			
Home > Organis	ns > Neisseria spp. > Neisseria isolates		
Neisser	<i>ia</i> isolates database		
	Query database	DI LOG IN	
	Search database		
	Browse, search by any criteria, or enter list of attributes.		+
		EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.	MALYSIS	
			+
		() INFORMATION	+
		E TYPING	

Jump to the 'Analysis' category, follow the link to 'Genome Comparator', then click 'Launch Genome Comparator'.

Genome Comparator

Summary: Compare genomes at defined loci or against loci defined in a reference genome

Genome Comparator 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. The analysis facilitates the determination of precisely which loci vary among isolates. Additionally, user-uploaded genomes that are not contained in the database can be analysed alongside genomes that are.

			544	567	663	140	447	662	670	671	672	655	61756	4178
Locus	Fullname	Product											(2837)	
NE150001 (pmC)		UDP 3-0 [3-tyd szonyrisztyl] N-sonykgiussownine deacetylase (JC 3.5.1.)												
NEIS0004		poptidy) prolyl siz-trans laamensee						×						
NEI50005		putative membrane protein												
NE(\$0006		of your allo dehoringenase												
NTI50007		methioryl CENA syntheticae												
NEISODDI		glacecomine-tructure-C-phosphote												
NE150011		hearthetical protein												
NE150015		UDP N analylgh assessing spropherspheryland												
Nb/50021		pytigtive inner membrione protein												
N8/50035		PUI-ble protoin												
NEISODSE		type IV pike retraction AliPees PK1												
NE/10041		25.5 transport competence and in	×	×			x				X			
N&/50041		putplice inner membrans transport protein.		1		6.5	6.0				_			
N6/50045			х	×	2	X	×		×	1	×			
NEI50046 (1524)		glucose 3-phoghate thymidplybram/erase												
NEI50047														
NEISCO4E (p+E)		UCP-glassee opresize	x	×	х	x	×	×	×		×			
NEIS0055	1405	capsule polysaccharide export outer membrane arguin	5	5	2	5	5	2	5	5	5			
NEIS0057	245	capsula polysaccharida expert inner membrane												
(195) NEISO258	0.0	protein captule polysactharide export ATP-binding protein												
(01775) NE150059		putative transmissional accessory protein												
NEI50059 NEI50062		tuncated gel5	×	×	×	â	÷.	×	x	Ű.	×		104	8

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.

			Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitidis	х	117			14	
	1		Germany	1999	carrier	Neisseria meningitidis	E	864				
	2		Germany	1999	carrier	Neisseria meningitidis	В	854	ST-18 complex			
	3		Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
5	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	Α	4	ST-4 complex	7	13-1	F1-5
	5		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
	6		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
	7		Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
	8		Germany	1999	carrier	Neisseria meningitidis	В	32	ST-32 complex			
5	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex	5-1	9	F3-1
	9		Germany	1999	carrier	Neisseria meningitidis	В	930	ST-334 complex			

ysis tool

🕒 Breakdown:	Fields	Two Field	Comb	binations	Polym	orphi	c sites	Publications	Sequ	ence bin
🛃 Analysis:	Codons	Gene Pres	sence	Genom	e Compara	ator	BLAST	rMLST spec	ies id	PCR
🐻 Export:	Dataset	Contigs	Sequ	ences						
📝 Third party:	GrapeTr	e iTOL	Phylo	Viz M	icroreact					

14.6.1 Analysis using defined loci

Select the isolate genomes that you wish to analyse. These will either be in a dropdown list or, if there are too many in the database, a text input where a list can be entered. You can also upload your own genomes for analysis - these should be either a single file in FASTA format (if you have just one genome), or a zip file containing multiple FASTA files. Select either the loci from the list or a set of schemes (either from the schemes box or from a list of recommended schemes if these have been set up). Press submit.

BIGSdb		
Home > Organisms > Neisseria spp. 3	Neisseria isolates > Plugins > Genome comp	arator
Genome Compara	tor - Neisseria isolates	Help 🔽 🚯
	efactual analysis because this database contains ed to select one of the recommended schemes.	different kinds of loci that are not equivalent (complete genes, MLST fragments, peptide variable regions,
		nake multiple selections in list boxes. In addition to selecting individual loci, you can choose to include all y, you can enter the accession number for an annotated reference genome and compare using the loci
Isolates	User genomes	Loci
644 ^ 662 665	Optionally include data not in the database.	165_rDNA - 165_rRNA (SSU_rRNA) 235_rRNA
666 667 669 -	Upload assembly FASTA file (or zip file containing multiple FASTA files - one per genome):	abcZ abcZ (NEIS1015) aceF (NEIS1279)
670 Zear List all	Choose file No file chosen	All None Paste list
Include in identifiers	Recommended sch	
isolate country region year date sampled isoyear sampled week sampled date received non culture	Select one or more sch below or use the full sc MLST Ribosomal MLST N. meningitidis cgMLS N. gonorrhoeae cgML	chemes list
Annotated reference genome	Parameters / options	Distance matrix calculation Alignments
Enter accession number: or choose annotated genome: or upload Genbank/EMBL/FASTA fi Choose file No file chosen	Min % identity: 70 Min % alignment: 50 BLASTN word size: 20	Completely exclude from analysis Completely exclude from analysis Treat as distinct allele dignore in pairwise comparison elignore in pairwise comparison Create alignment stats
Core genome analysis	Filter by	Action
Core threshold (%): 90 🗸 👔	Sequence method: Project: Experiment:	

The job will be submitted to the job queue and will start running shortly.

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.

	ver							
Submit time: Status: Start time: Progress:	2020-07-31 09:07:4	59 +2	-					
Output								
Analysis against define	1 loci							
All loci								
Allele numbers are used whe are marked as 11	re these have been d	efined, otherwise s	equences will be mark	ed as 'New#1, 'New#2' et	c. Missing alleles ar	e marked as X'.	Incomplete alleles (locate	ed at end of contig)
Locus Full name Product 6 abcZ	2 2	2	666 (2843) 667 (2842) 2 2 2		2 2	2	41784 (2839) 41785 (28 2 2	338)
adk aroE fumC	3 3 4 1 3 3	9 4	3 3 4 4 23 23		3 3 19 19 3 3			
gdh pdhC	8 8 4 4	8 4 4	8 8 6 6	8 8 4 4	8 8 4 4	8 4	8 8 4 4	
pgm Loci with sequence diff	6 6		6 6	6 6	6 6	6	6 6	-
Variable loci: 4	erences among is	olates						
Locus Full name Product 6 abcZ						847) 698 (FAM18)	41784 (2839) 41785 (28	338)
aroE fumC	2 2 4 1 3 3	9 4	2 2 4 4 23 23	7 2 4 19 3 3	2 2 19 19 3 3			
pdhC	4 4	4	6 6	4 4	4 4	4	4 4	-
Exactly matching loci These loci are identical in all	isolates							
Matches: 3								
Locus Full name Product 6 adk gdh ggm ggm ggm ggm ggm ggm ggm ggm ggm gg	44 (L93/4286) 662 (2 3 3 8 8 6 6		666 (2843) 667 (2842) 3 3 8 8 6 6	669 (2846) 670 (2840) 3 3 8 8 6 6	671 (2844) 672 (2 3 3 8 8 6 6	3	41784 (2839) 41785 (28 3 3 8 8 6 6	338)
Unique strains								
Unique strains: 5								
onique sciains. 5	Strain 3 Strain 666 (2843) 669 (2842) 667 (2842) 669							
Strain 1 Strain 2 662 (2837) 644 (L93/4286) 670 (2840) 698 (FAM18) 671 (2844) 41784 (2839)								
Strain 1 Strain 2 662 (2837) 644 (L93/4286) 670 (2840) 698 (FAM18) 671 (2844) 41784 (2839) 672 (2847) 41785 (2838)								
Strain 1 Strain 2 662 (2837) 644 (L93/4286) 670 (2840) 698 (FAM18) 671 (2844) 41784 (2839)								
Strain 1 Strain 2 662 (2837) 644 (L93/4286) 670 (2840) 698 (FAM18) 671 (2844) 41784 (2839) 672 (2847) 41785 (2838) Charts Charts	equency							
Strain 1 Strain 2 662 (2837) 644 (L95)/4286) 670 (2840) 698 (FAM18) 670 (2840) 698 (FAM18) 671 (2844) 41784 (2839) 672 (2847) 41785 (2838) 672 (2847) 41785 (2838) Charts Click to enlarge. Locus presence fr D	and the second s							
Strain 1 Strain 2 662 (2837) 644 (L93/4286) 670 (2840) 698 (FAM18) 671 (2844) 41784 (2839) 672 (2847) 41785 (2838) Charts Click to enlarge. Locus presence fr 9 9 9 9 9 9 9 9 9 9 9 9 9	and the second s							
Strain 1 Strain 2 662 (2837) 644 (L93/4286) 670 (2840) 698 (FAM18) 671 (2844) 41785 (2838) Charts Click to enlarge. Click to enlarge. Click to enlarge.	and the second s	Ex Ex	cel format output		E	Splits graph (Neig SVG format)	ghbour-net;	

14.6.2 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.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Genome con	nparator
Genome Comparator - Neisseria isolate	S
Selecting 'All loci' will produce an artefactual analysis because this database contair intergenic regions etc.). You are advised to select one of the recommended schemes	ns different kinds of loci that are not equivalent (complete genes, MLST fragments, peptide variable regions, ;.
	o make multiple selections in list boxes. In addition to selecting individual loci, you can choose to include all rely, you can enter the accession number for an annotated reference genome and compare using the loci
Isolates User genomes	Include in identifiers
644 Optionally include data not in the database. 665 Upload assembly FASTA file 666 (or zip file containing multiple 669 FASTA files - one per genome): () 670 Choose file	Isolate country region year date sampled Isoyear sampled week sampled date received non culture
Annotated reference genome — Parameters / options –	Distance matrix calculation Alignments
Enter accession number: Min % identity: 7 Image: Construction of the second	50 Completely exclude from analysis Include ref sequences in alignment 20 Image: Stinct allele Align all loci (not only variable) Create alignment stars Create alignment stars
Core genome analysis Filter by	Action
Core threshold (%): 90 V () Calculate mean distances () Experiment: V ()	▼ ③ ▼ ③ RESET SUBMIT

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.

analysis by reference	e genome															
access	ion: AM421808															
vers	ion: 1															
t	/pe: dna															
	ath: 2194961															
descript		M19 complete														
		MIS complete	e genome.													
coding regi	ons: 1975															
ch unique allele is def	ined a number starting at 1. Missing alleles	are marked as	'X' . Inco	omplete all	eles (locateo	d at end	of cont	ig) are n	narked a	is T.						
h unique allele is def	ined a number starting at 1. Missing alleles Product			Reference		662	665	666	667	669	670 (2840)	671 (2844)	672 (2847)	698 (FAM18)	41784	
Locus		Sequence	Genome	Reference	644	662	665	666	667	669						
Locus 4C0001 lpxC envA	Product UDP-3-O-[3-hydroxymyristoyl] N-	Sequence length	Genome position	Reference	644	662	665	666	667	669						
Locus IC0001 lpxC envA NMC0002 pilS1	Product UDP-3-O-[3-hydroxymyristoyl] N- acetylglucosmine deacetylase	Sequence length 924	Genome position 1261	Reference genome 1	644	662	665	666 (2843) 1	667 (2842) 1	669	(2840) 1	(2844) 1		(FAM18) 1	(2839) 1	
Locus MC0001 lpxC envA NMC0002 pil51	Product UDP-3-0-[3-hydroxymyristoyl] N- acetylglucosmine deacetylase pilin (fragment)	Sequence length 924 291	Genome position 1261 3341	Reference genome 1	644 (L93/4286) 1	662 (2837) 1	665 (2845) 1 1	666 (2843) 1	667 (2842) 1	669 (2846) 1 1	(2840) 1 1	(2844) 1 1	(2847) 1 1	(FAM18) 1 1	(2839) 1 2	
Locus IC0001 lpxClenvA NMC0002 pilS1 NMC0003 pilS2 NMC0005	Product UDP-3-0-[3-hydroxymyristoyl] N- acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-prolyl cis-trans isomerase putative membrane protein	Sequence length 924 291 366	Genome position 1261 3341 3675 4069 4476	Reference genome 1	644 (L93/4286) 1 1 2	662 (2837) 1 1 2	665 (2845) 1 1 2	666 (2843) 1 1 2	667 (2842) 1 1 2	669 (2846) 1 1	(2840) 1 1 2	(2844) 1 1 2	(2847) 1 1 2	(FAM18) 1 1	(2839) 1 2 2 2 3	
Locus MC0001 lpxClenvA NMC0002 pilS1 NMC0003 pilS2 NMC0004 fbp	Product UDP-3-O-[3-hydroxymyristoyl] N- acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-prolyl cis-trans isomerase	Sequence length 924 291 366 330	Genome position 1261 3341 3675 4069	Reference genome 1 1 1 1	644 (L93/4286) 1 2 2	662 (2837) 1 1 2 2	665 (2845) 1 1 2 2	666 (2843) 1 1 2 2	667 (2842) 1 1 2 2	669 (2846) 1 1	(2840) 1 2 2	(2844) 1 2 2	(2847) 1 2 2	(FAM18) 1 1 1 1	(2839) 1 2 2 2	
Locus MC0001 lpxClenvA NMC0002 pilS1 NMC0003 pilS2 NMC0004 fbp NMC0005 NMC0006	Product UDP-3-O-[3-hydroxymyristoyl] N- acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-prolyl cis-trans isomerase putative membrane protein putative glycerate dehydrogenase methionyl-tRNA synthetase	Sequence length 924 291 366 330 219	Genome position 1261 3341 3675 4069 4476 4816 5843	Reference genome 1 1 1 1	644 (L93/4286) 1 2 2 2 2 2 2 2 2 2 2	662 (2837) 1 2 2 3	665 (2845) 1 2 2 3	666 (2843) 1 2 2 3 2 2 2	667 (2842) 1 2 2 3	669 (2846) 1 2 1 4	(2840) 1 2 2 3	(2844) 1 2 2 5	(2847) 1 2 2 3 2 2 2 2 2	(FAM18) 1 1 1 1	(2839) 1 2 2 2 3 2 3 2 3 3	
Locus MC0001 lpxClenvA NMC0002 pilS1 NMC0003 pilS2 NMC0004 fbp NMC0005	Product UDP-3-O-[3-hydroxymyristoyl] N- acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-prolyl cis-trans isomerase putative membrane protein putative giycerate 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 2	665 (2845) 1 2 2 3 2 2	666 (2843) 1 2 2 3 2 2	667 (2842) 1 2 2 3 2	669 (2846) 1 2 1 2 1 4 2	(2840) 1 2 2 3 2 2	(2844) 1 2 2 5 2	(2847) 1 2 2 3 2 2	(FAM18) 1 1 1 1 1 1 1	(2839) 1 2 2 2 3 2 3 2	

14.6.3 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	4
country	
region	
year	
date sampled	
isoyear sampled	
week sampled	
date received	
non culture	Ŧ

Multiple values can be selected by clicking while holding down Ctrl.

14.6.4 Reference genome fieldset

This section allows you to choose a reference genome to use as the source of comparator sequences.

Annotated reference genome	
Enter accession number:	
i	
or choose annotated genome:	
FAM18 (Nm) 🗸	
or upload Genbank/EMBL/FASTA file:	
Choose file No file chosen	(i)
	\sim

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.

14.6.5 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: 20 🗸 👔
Rescan undesignated loci 🧃

- 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.

14.6.6 Distance matrix calculation fieldset

This section provides options for the treatment of incomplete and paralogous loci when generating the distance matrix.

Distance matrix calculation –

With incomplete loci:

- Completely exclude from analysis
- Treat as distinct allele
- Ignore in pairwise comparison
- Exclude pairwise missing loci
- Exclude loci paralogous in all
- Exclude pairwise paralogous loci

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.

14.6.7 Alignments fieldset

This section enables you to choose to produce alignments of the sequences identified.

Alignments
Produce alignments (i)
Include ref sequences in alignment
Align all loci (not only variable)
Create alignment stats
Aligner: MAFFT 🗸

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 longer 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.

14.6.8 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 🖌 👔
Calculate mean distances 👔

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.

14.6.9 Filter fieldset

This section allows you to further filter your collection of isolates and the contigs to include.

Filter by			
Sequence method:		* (i)	
Project:		-	~ (i)
Experiment:	~ (i)		Ũ

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.

14.6.10 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. nce matrix calculated based on allelic differences and a NeighborNet graph generated from this distance matrix.

14.7 GrapeTree

GrapeTree is a tool for generating and visualising minimum spanning trees. It has been developed to handle large datasets (in the region of 1000s of genomes) and works with 1000s of loci as used in cgMLST. It uses an improved minimum spanning algorithm that is better able to handle missing data than alternative algorithms and is able to produce publication quality outputs. Datasets can include metadata which allows nodes in the resultant tree to be coloured interactively.

GrapeTree can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates	
Neisseria isolates database	
Query database	DILOG IN
Search database	
Browse, search by any criteria, or enter list of attributes.	PROJECTS +
	EXPORT +
Search by combinations of loci This can include partial matches to find related isolates.	MALYSIS
This can include partial matches to find related isolates.	CUSTOMISE +
	INFORMATION +
	S TYPING

Jump to the 'Third party' category, follow the link to GrapeTree, then click 'Launch GrapeTree'.

GrapeTree can be accessed from the contents page by clicking the 'GrapeTree' link.

GrapeTree
Summary: Visualization of genomic relationships
GrapeTree is a tool for generating and visualising minimum spanning trees (Zhou at al. 2018 Genome Res 28:1395-1404). It has been developed to handle large datasets (in the region of 1000s of genomes) and works with 1000s of loci as used in cgMLST. It uses an improved minimum spanning algorithm that is better able to handle missing data than alternative algorithms and is able to produce publication quality outputs. Datasets can include metadata which allows nodes in the resultant tree to be coloured interactively.
Documentation bigsdbreadthedocs.io 😰
Launch 'GrapeTree'

Alternatively, it can be accessed following a query by clicking the 'GrapeTree' button at the bottom of the results table. Isolates returned from the query will be automatically selected within the GrapeTree interface.

19958 M10 240476	UK [England] 2010 invasive (unspec	fied/other) Neisseria meningitidis	В	9812 ST-213 complex	22	14	F5-5
19959 M10 240499	UK [England] 2010 invasive (unspec	fied/other) Neisseria meningitidis	В	275 ST-269 complex	22	9	F5-12
19960 M10 240500	UK [England] 2010 invasive (unspec	fied/other) Neisseria meningitidis	В	41 ST-41/44 complex	7-2	4	F1-5
19961 M10 240502	UK [England] 2010 invasive (unspec	fied/other) Neisseria meningitidis	В	340 ST-41/44 complex	7-2	4	F1-5
19962 M10 240503	UK [England] 2010 invasive (unspec	ified/other) Neisseria meningitidis	Y	23 ST-23 complex	5-1	2-2	F1-96
19963 M10 240505	UK [England] 2010 invasive (unspec	ified/other) Neisseria meningitidis	Y	1655 ST-23 complex	5-1	10-10	F4-1
19964 M10 240507	UK [England] 2010 invasive (unspec	ified/other) Neisseria meningitidis	Y	183 ST-23 complex	21	16-5	deleted
19965 M10 240508	UK [England] 2010 invasive (unspec	fied/other) Neisseria meningitidis	В	1575	7-2	13-1	F1-7
19966 M10 240511	UK [England] 2010 invasive (unspec	fied/other) Neisseria meningitidis	В	4713	22	9	F5-12
19967 M10 240512	UK [England] 2010 invasive (unspec	fied/other) Neisseria meningitidis	В	41 ST-41/44 complex	7-1	1	F1-5
19968 M10 240514	UK [England] 2010 invasive (unspec	ified/other) Neisseria meningitidis	W	11 ST-11 complex	5	2	F1-1
19969 M10 240515	UK [England] 2010 invasive (unspec	ified/other) Neisseria meningitidis	В	269 ST-269 complex	5-1	2-2	F5-1
19970 M10 240520	UK [England] 2010 invasive (unspec	ified/other) Neisseria meningitidis	Y	1655 ST-23 complex	5-1	10-1	F4-1
\bigcirc \bigcirc 1							
(YC) 1 2							

Select the isolates to include. The tree can be generated from allelic profiles of any selection of loci, or more conveniently, you can select a scheme in the scheme selector, or choose from recommended schemes if these have been set, to include all loci belonging to that scheme.

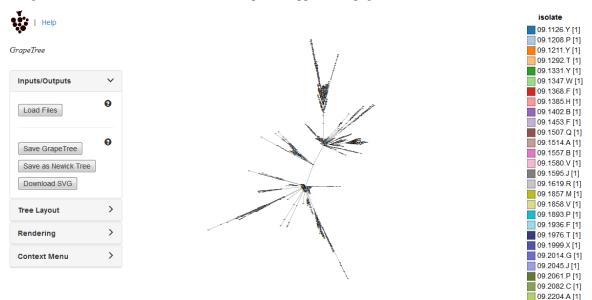
Additional fields can be selected to be included as metadata for use in colouring nodes - select any fields you wish to include. Multiple selections can be made by holding down shift or ctrl while selecting. Click 'Submit' to start the analysis.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > GrapeTree
GrapeTree: Visualization of genomic relationships - Neisseria isolates
This plugin generates a minimum-spanning tree and visualizes within GrapeTree: GrapeTree is developed by: Zhemin Zhou (1), Nabil-Fareed Alikhan (1), Martin J. Sergeant (1), Nina Luhmann (1), Cátia Vaz (2,5), Alexandre P. Francisco (2,4), João André Carriço (3) and Mark Achtman (1) 1. Warwick Medical School, University of Warwick, UK 2. Instituto de Engenharia de Sistemas e Computadores: Investigação e Desenvolvimento (INESC-ID), Lisboa, Portugal 3. Universidade de Lisboa, Faculdade de Medicina, Instituto de Microbiologia and Instituto de Medicina Molecular, Lisboa, Portugal 5. ADEETC, Instituto Superior Técnico, Universidade de Lisboa, Instituto Politécnico de Lisboa, Portugal 5. ADEETC, Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Portugal Publication: Zhou at al. (2018) GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res 28:1395-1404.
Selecting All loci' will produce an artefactual analysis because this database contains different kinds of loci that are not equivalent (complete genes, MLST fragments, peptide variable regions, intergenic regions etc.). You are advised to select one of the recommended schemes. This tool will generate minimum spanning trees trees from allelic profiles. Please check the loci that you would like to include. Alternatively select one or more schemes to include all loci that are members of the scheme. Analysis is limited to 50,000 records.
Isolates Loci Recommended schemes Schemes 18968 165_rDNA Select one or more schemes below or use the full schemes list. 19023 235_rRNA below or use the full schemes list. 19024 abcZ abcZ 19025 abcZ N. meningitidis cgMLST v1.0 10026 Clear Atl. None Clear List atl. Atl. None
Include fields Parameters / options Action Select additional fields to include in GrapeTree metadata. Ciprofloxacin range Cional complex (MLST) Cluster (GC OMV peptide typing) Comments Continent Country

The job will be sent to the job queue. When it has finished, click the button marked 'Launch GrapeTree'.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Job status viewer
Job status viewer
Status Job id: BIGSdb_001935_1596193067_67514 Submit time: 2020-07-31 10:57:47 Status: Finished Start time: 2020-07-31 10:57:50 Progress: 100% Stop time: 2020-07-31 11:19:38 Total time: 21 minutes and 47 seconds
Output Launch GrapeTree Files
Profiles (TSV format) (18.7 MB)
Tar file containing all output files (only files <10MB included - download larger files separately)
Please note that job results will remain on the server for 7 days.

The generated tree will be rendered in the GrapeTree application page.

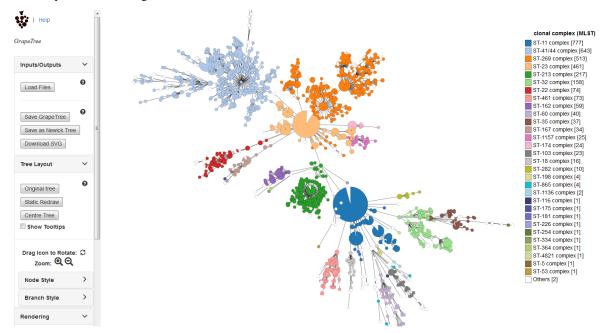


The image can be manipulated in various ways. These include modifying the tree layout, customising node labels and size, modifying branch lengths and collapsing branches. The image can be saved in SVG format which can be further edited in image publishing software such as Inkscape.

As an example, the default cgMLST tree (above) has been modified (below) as follows:

- Nodes coloured by clonal complex
- Labels removed

- Branches collapsed where <=100 loci different
- Node size set to 200%
- Kurtosis (node size relative to number of isolates) set to 75%
- Dynamic rendering allowed to run to fan out nodes



Full details can be found in the GrapeTree manual.

Note: GrapeTree has been described in the following publication:

Z Zhou, NF Alikhan, MJ Sergeant, N Luhmann, C Vaz, AP Francisco, JA Carrico, M Achtman (2018) GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res 28:1395-1404.

14.8 In silico PCR

This is a tool that can be used to simulate PCR reactions run against genomes stored in the database. This is useful for designing and testing primers. The plugin uses the exonerate ipcress program to perform its simulation.

The tool can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb						
Home > Organisms > Neisseria spp. > Neisseria isolates						
Neisseria isolates database						
Query database	D LOG IN					
Search database						
Browse, search by any criteria, or enter list of attributes.	PROJECTS	+				
	Z EXPORT	+				
Search by combinations of loci This can include partial matches to find related isolates.	MALYSIS					
		+				
		+				
	S TYPING					

Jump to the 'Analysis' category, follow the link to BLAST, then click 'Launch PCR'.

PCR
Summary: In silico PCR tool for designing and testing primers
The <i>in silico</i> PCR tool can be used to simulate PCR reactions run against genomes stored in the database. This is useful for designing and testing primers. The plugin uses the exonerate ipcress program to perform its simulation. Primers can contain wobble bases and the number of allowed mismatches for each primer can be set independently. The output is a table of predicted products, showing the number of products and their positions within a contig. It is also possible to export the predicted product sequence.
Documentation bigsdb.readthedocs.io 🗗
Corput Image: An order of the start of the s
Launch 'PCR'

Alternatively, it can be accessed following a query by clicking the 'PCR' button at the bottom of the results table. Isolates returned from the query will be automatically selected within the analysis interface.

/	1971	B34; NIBSC_2/00; 21034	Finiand	13/2	invasive (unspecified/other)	iveisseria meningiliuis	А	С	51-5 complex	20	У	F0-1
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	18-1	3	F5-1
11	129E	B92; NIBSC_2828; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F3-6
13	139M	B99; NIBSC_2795; Z1099	Philippines	1968		Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
24	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	5-1	9	F3-1
30	14	BennettTree10; alpha14	Germany	1999	carrier	Neisseria meningitidis	cnl	53	ST-53 complex	7	30-3	F5-5
31	10	B269; NIBSC_2825; Z1269	Burkina Faso	1963	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
34	20	B275; NIBSC_2767; Z1275	Niger	1963	invasive (unspecified/other)	Neisseria meningitidis	A			5-2	10	F1-7
35	26	B278; NIBSC_2764; Z1278	Niger	1963	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13	F1-5
52	243	B362; NIBSC_2779; Z1362	Cameroon	1966		Neisseria meningitidis	A	4	ST-4 complex	7	13	F1-5
61	393	B392; NIBSC_2823; Z1392	Greece	1968	carrier	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
64	254	B439; NIBSC_2812; Z1439	Djibouti	1966	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F1-7
67	S5611	B466; NIBSC_2765; Z1466	Australia	1977	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
82	11-004	B503; NIBSC_2826; Z1503	China	1984	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-8
84	IAL2229	B506; NIBSC_2816; Z1506	Brazil	1976		Neisseria meningitidis	A	5	ST-5 complex	20	9	F2-1
90	CN100	B534; NIBSC_2768; Z1534	UK [England]	1941	invasive (unspecified/other)	Neisseria meningitidis	A	21		5-2	10	F3-9

Analysis tools

Breakdown: Fields Two Field Combinations Polymorphic sites Publications Sequence bin

Analysis: BURST Codons Gene Presence Genome Comparator BLAST rMLST species id PCR

Export: Dataset Contigs Sequences

Third party: GrapeTree iTOL PhyloViz Microreact

Select the isolates to include. These will be pre-populated if you arrive here following a search.

Enter your forward and reverse primer sequences in the appropriate boxes. These may contain wobble bases if necessary. You can also specify how many mismatches are allowed for each primer. Finally, you can restrict the reported length to only those products that fall between a minimum and maximum length range.

BIGSdb								
Home > Organisms > Neiss	seria spp. > Neisseria isolates > Plugins > Ir	n silico PCR						
In silico PCR	In silico PCR							
Use this tool to simulate P written by Guy Slater.	CR reactions run using genomes stored in the	e database. This is useful for designing	and testing primers. The plugin is a v	vrapper for the exonerate ipcre	ess program			
Please select the required	isolate ids to run the PCR reaction against. Th	nese isolate records must include gene	ome sequences.					
Isolates	Primer 1	Primer 2	Reported products	Action				
1 2 7 10 11 13 19	Allowed mismatches: 0 V	TCCCCGTCGTAAAAAACAATC	Min length: 0 Max length: 10000 □ Export sequences	SUBMIT				
Clear List all								

Click 'Submit'. The job will be sent to the job queue. The output will be a table of predicted products, showing the number of products and their positions within a contig. A summary of this table is also available to download in tabdelimited text of Excel formats.

bit Status viewer Distatus viewer Status Lob id: BIGSdb_028633_1596197976_78216 Status: Status: Status: Status: Status: Finished: Status: 2020-07-31 12:19:39 Progress: 100% Stop time: 2020-07-31 12:19:49: Stop time: 2020-07-31 12:19:41 Total time: 2 seconds Dutput 1 1 MAMM027 1 180675 1 180675 1 180675 1 180675 1 180675 1 180675 1 180675 1 180755 1 181206 1 181206 1 181205 1 181205 1 180675 1 181206 1 181206 1 181208 1 181208 1 181208 1 181208	
<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	
<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	
Submit tim:: 2020-07-31 12:19:39 Progres: 100% Start tim:: 2020-07-31 12:19:39 Depres: 2020-07-31 12:19:40 Total tim:: 2 seconds	
Start tim:: 2020-07-31 12:19:41. Gramming:: 2020-07-31 12:19:41. Tat tim:: 2 seconds	
Stop time: 2020-07-31 12:19:41 Total time: 2 seconds Dutput Image: Seconds Image: Seconds Image: Seconds Image: Seconds Dutput Image: Seconds	
Dutput 1 10012 1 180126 898 6772 7669 forward 2 12014 1 180075 898 6772 7669 forward 2 12014 1 180075 898 6772 7669 forward 2 12014 1 180075 898 6707 7667 forward 2 12014 1 18075 898 6723 7667 forward 1 129E 1 181705 898 1623 2517 reverse 13 139M 1 181716 898 6723 7687 forward 10 1 182398 898 1632 2517 reverse 13 139M 1 18216 898 6747 7617 forward 10 1 182398 898 1520 2517 reverse 755611 1 183505 898 1624 2511 reverse 12 120 1 18352 898	
Solate PCR +ve products contig length stat end description 1 A4/M1027 1 180126 898 6772 7669 forward 2 120M 1 180075 898 f108 2505 reverse 7 7891 1 180075 898 f108 2505 reverse 10 6748 2 181299 898 f1614 2511 reverse 11 129E 1 181720 898 f623 2520 reverse 13 139M 1 18216 898 6700 7687 forward 19 53131 1 1122556 898 6724 7611 forward 30 14 1 8 898 98430 984327 forward 31 10 1 182398 898 1624 2511 reverse 52 243 1 18357	
Solate PCR +ve products contig length stat end description 1 A4/M1027 1 180126 898 6772 7669 forward 2 120M 1 180075 898 f108 2505 reverse 7 7891 1 180075 898 f108 2505 reverse 10 6748 2 181299 898 f1614 2511 reverse 11 129E 1 181720 898 f623 2520 reverse 13 139M 1 18216 898 6700 7687 forward 19 53131 1 1122556 898 6724 7611 forward 30 14 1 8 898 98430 984327 forward 31 10 1 182398 898 1624 2511 reverse 52 243 1 18357	
I A4/M1027 I 180126 B98 6772 7669 forward 1 A4/M1027 I 180126 898 6772 7669 forward 2 120M I 180075 898 1608 2505 reverse 7 7891 I 180795 898 6707 7667 forward 10 6748 2 181299 898 1614 2511 reverse 11 129E I 181570 898 1623 2520 reverse 13 139M I 18216 898 6730 7687 forward 10 Y 1 182205 898 1624 2517 reverse 13 139M I 182216 898 6737 forward 10 Y 1 182398 898 1842 2517 reverse 13 19M I 182398 898 184	
1 A4/M1027 1 180126 898 6772 7669 forward 2 120M 1 180675 898 1608 2505 reverse 7 7891 1 18075 898 1614 2511 reverse 10 6748 2 181209 898 1614 2511 reverse 11 129E 1 181720 898 1620 2517 reverse 13 139M 1 181216 898 6727 forward 19 53131 1 181205 898 1620 2517 reverse 30 14 1 8298 1620 2517 reverse 52 26 1 18239 898 1620 2517 reverse 53 10 1 18239 898 1620 2517 reverse 53 26 1 18279 898 1620 2517 reverse 54 20 1 182398 898 1620	
2 120M ✓ 1 180675 898 1608 2505 reverse 7 7891 ✓ 1 180795 898 670 7667 forward 0 6748 ✓ 2 181299 898 1614 2511 reverse 181505 898 1623 2520 reverse 181520 898 1623 2517 reverse 13 139M ✓ 1 18216 898 670 7667 forward 19 53131 ✓ 1 18225 898 1620 2517 reverse 10 ✓ 1 182216 898 1620 2517 reverse 13 130 ✓ 1 18225 898 1620 2517 reverse 14 ✓ 1 8 898 98430 984327 forward 15 10 ✓ 1 182398 898 1610 2517 reverse 26 ✓ 1 183555 898	
10 6748 2 181299 898 1614 2510 reverse 181505 898 1623 2520 reverse 181570 898 1623 2520 reverse 181570 898 1623 2517 reverse 181720 898 182225 898 182225 898 18225 898 6774 7617 forward 11 11 1215556 898 6774 7611 forward 10 1 11 11 11 121556 898 98430 984327 forward 513 10 1 18239 898 18210 reverse 52 24 1 18352 898 1612 18351 898 1612 18351 18453 18453 18453 1988 198 198 198 198 198	
11 129E 139E 1311 139M 1 181216 898 1620 2517 reverse 1812125 898 1620 2517 reverse 1812125 898 1620 2517 reverse 1 182225 898 1215556 898 6774 7671 forward 1 1	
19 S3131 1 182225 898 1620 2517 reverse reverse	
30 14 ✓ 1 8 988 983430 984327 forward 11 10 ✓ 1 182398 898 1620 2517 reverse 45 20 ✓ 1 182791 898 6982 7879 forward 35 26 ✓ 1 183228 898 1624 2521 reverse 52 243 ✓ 1 183555 898 6767 7664 forward 54 254 ✓ 1 183761 898 1616 2513 reverse 57 55611 ✓ 1 184376 898 1619 2516 reverse 52 11-004 ✓ 1 184525 898 6770 7667 forward 52 11-004 ✓ 1 18457 898 1505 forward 64 222.9 ✓ 1 18457 898 7208 8105 forward 90 CN100 ✓ 1 184677 <	
34 20 ✓ 1 182791 898 6982 7879 forward 35 26 ✓ 1 182948 898 1913 2810 reverse 22 243 ✓ 1 183522 898 1624 2521 reverse 51 393 ✓ 1 183555 898 6767 7664 forward 64 254 ✓ 1 183555 898 1612 2513 reverse 64 254 ✓ 1 184575 898 1619 2516 reverse 82 11-004 ✓ 1 184377 898 1911 2808 reverse 82 11-004 ✓ 1 18457 898 7067 forward 90 CN100 ✓ 1 184677 898 7208 8105 forward Pilles Text format summary file Excel format summary file FASTA file	
52 243 ✓ 1 183322 898 1624 2521 reverse 61 393 ✓ 1 183555 898 6767 7664 forward 64 254 ✓ 1 183565 898 1616 2513 reverse 65 55611 ✓ 1 184797 898 1619 2516 reverse 82 11-004 ✓ 1 184378 898 1911 2808 reverse 84 IAL2229 ✓ 1 184577 898 1911 2808 forward 90 CN100 ✓ 1 184577 898 700 7667 forward Files Excel format summary file Excel format summary file FASTA file	
54 254 ✓ 1 183761 898 1616 2513 reverse 57 55611 ✓ 1 184197 898 1619 2516 reverse 2 11-004 ✓ 1 184357 898 1911 2808 reverse 34 IAL2229 ✓ 1 184525 898 6770 7667 forward 90 CN100 ✓ 1 184677 898 7208 8105 forward Viles 1 84677 898 7208 forward	
82 11-004 ✓ 1 184337 898 1911 2808 reverse 84 JAL2229 ✓ 1 184525 898 6770 7667 forward 90 CN100 ✓ 1 184677 898 7208 8105 forward Piles Fast format summary file Fast format summary file FASTA file	
84 IAL2229 1 184525 898 6770 7667 forward 90 CN100 1 184677 898 7208 8105 forward Files Fast format summary file Excel format summary file FASTA file	
Files	
Text format summary file	
Text format summary file Excel format summary file FASTA file product se	
E CACE format summary nee	of predicted
Tar file containing all output	
files	

It is also possible to export the predicted product sequence. You can do this by selecting the 'Export sequences' checkbox on the options form.

Reported products						
Min length:	0					
Max length:	10000					
Export sequences						

Note: The exported sequences will include the primer regions. It is important to note that, unlike a real PCR reaction, these sequences represent the sequence within this region in the genome. In a real PCR reaction, the primers are themseleves incorporated in to the product, so even if there was a mismatch in the primer region, the product sequence

would include the primer sequence.

14.9 Interactive Tree of Life (iTOL)

The ITOL plugin allows you to generate and visualise phylogenetic trees calculated from concatenated sequence alignments of selected loci (or the loci belonging to a particular scheme). Currently, only Neighbour-joining trees are supported. Datasets can include metadata which allows nodes in the resultant tree to be coloured.

ITOL can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb								
Home > Organis	Home > Organisms > Neisseria spp. > Neisseria isolates							
Neisser	Neisseria isolates database							
	Query database		→J LOG IN					
	Search database							
	Browse, search by any criteria, or enter list of attributes.			+				
			EXPORT	+				
	Search by combinations of loci This can include partial matches to find related isolates.		🗠 ANALYSIS					
			CUSTOMISE	+				
			1 INFORMATION	+				
			S TYPING					

Jump to the 'Third party' category, follow the link to iTOL, then click 'Launch iTOL'.

ITOL can be accessed from the contents page by clicking the 'iTOL' link.

iTOL
Summary: Phylogenetic trees with data overlays
The ITOL plugin allows you to generate and visualise phylogenetic trees calculated from concatenated sequence alignments of selected loci (or the loci belonging to a particular scheme). Currently, only Neighbour-joining trees are supported. Datasets can include metadata which allows nodes in the resultant tree to be coloured. Datasets are uploaded to the Interactive Tree of Life website (Letunic & Bork 2016 Nucleic Acids Res 44(W1):W242-5) for visualisation.
Documentation bigsdbreadthedocs io 🗹
Launch 'ITOL'

Alternatively, it can be accessed following a query by clicking the 'iTOL' button at the bottom of the results table. Isolates returned from the query will be automatically selected within the iTOL interface.

F1-
F1-
F1-
F1-
F1-
1.7.
F3-

G Breakdown:	Fields	Two Field	Combinations	Polymorphic sites	Publications	Sequence bin	
🛃 Analysis:	BURST	Codons	Gene Presence	Genome Comparato	r BLAST	rMLST species id	PCR
B Export:	Dataset	Contigs	Sequences				
🛃 Third party:	GrapeTre	e iTOL	PhyloViz M	icroreact			

Select the isolates to include. The tree can be generated from concatenated sequences of any selection of loci, or more conveniently, you can select a scheme in the scheme selector, or in the list of recommended schemes if these have been set up, to include all loci belonging to that scheme.

Additional fields can be selected to be included as metadata for use in colouring nodes - select any fields you wish to include in the 'iTOL datasets' list. Multiple selections can be made by holding down Shift or Ctrl while selecting. You can also choose how nodes are labeled by metadata - either by colouring the labels or using coloured strips.

Click 'Submit' to start the analysis.

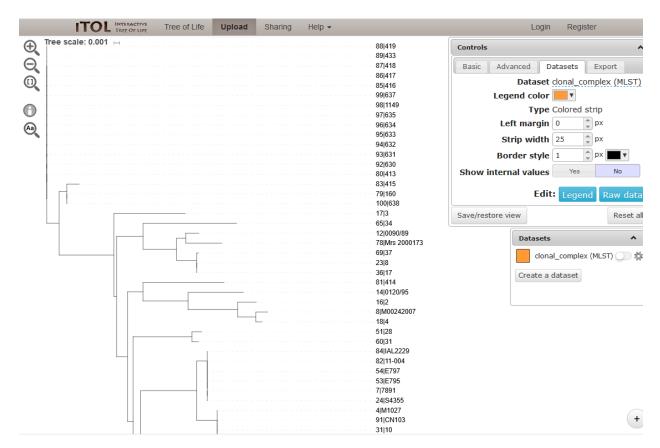
BIGSdb									
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > iTOL									
iTOL - Interactive Tree of Life - Neisseria isolates									
ITOL is develop 1. Biobyte s 2. European 3. Max Delb 4. Department Web site: https://toLembLde/	This plugin uploads data for analysis within the Interactive Tree of Life online service: iTOL is developed by: Ivica Letunic (1) and Peer Bork (2,3,4) 1. Biobyte solutions GmbH, Bothestr 142, 69126 Heidelberg, Germany 2. European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heideberg, Germany 3. Max Delbrück Centre for Molecular Medicine, Silts Perint, Germany 4. Department of Bioinformatics, Biocenter, University of Würzburg, 97074 Würzburg, Germany 4. Department of Bioinformatics, Biocenter, University of Würzburg, 97074 Würzburg, Germany Web site: https://toLembLde/ Publication: Letunic & Bork (2016) Interactive tree of life (ITOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44(W1):W242-5.								
peptide loci with genome sequence scheme.	This tool will generate neighbor-joining trees from concatenated nucleotide sequences. Only DNA loci that have a corresponding database containing allele sequence identifiers, or DNA and peptide loci with genome sequences, can be included. Please check the loci that you would like to include. Alternatively select one or more schemes to include all loci that are members of the scheme. Analysis is limited to 2,500 records or 110,000 sequences (records x loci).								
I I I I I I I I I I I I I I I I I I I	User genomes Optionally include data not in the database. Upload assembly FASTA file (or zip file containing multiple FASTA files - one per genome): () Choose file No file chosen	Loci '165_rDNA 165_rRNA (SSU_rRNA) 22S_rRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) All None Paste list	Recommended schemes Select one or more schemes below or use the full schemes list. MLST Ribosomal MLST N. meningitidis cgMLST v1.0 N. gonorrhoeae cgMLST v1.0						
Schemes Plasmids Typing WLST Finetyping ant Schemes	azithromycin spectinomycin bioproject accession biosample accession	iTOL data type							

The job will be sent to the job queue. When it has finished, the generated tree and associated metadata will be uploaded

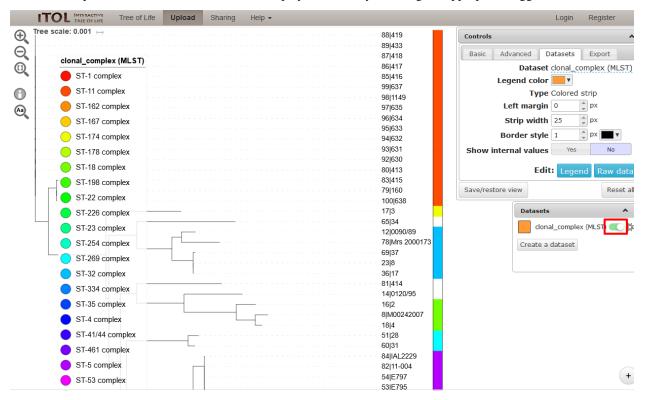
to the Interactive Tree of Life website (https://itol.embl.de/). Click the button marked 'Launch iTOL'.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Job status viewer
Job status viewer
Status Job Ids BIGSdb_027908_1596206761_04985 Submit time 2020-07-31 14:46:01 Status finished Status 2020-07-31 14:46:05 Progress 2020-07-31 14:46:05 Stop time 2020-07-31 14:46:05 Progress 2000-07-31 14:46:05 Dots 2000-07-31 14:46:25 Stop time 20 seconds
Output Launch ITOL Files
Concatenated FASTA
Tar file containing all output files
Please note that job results will remain on the server for 7 days.

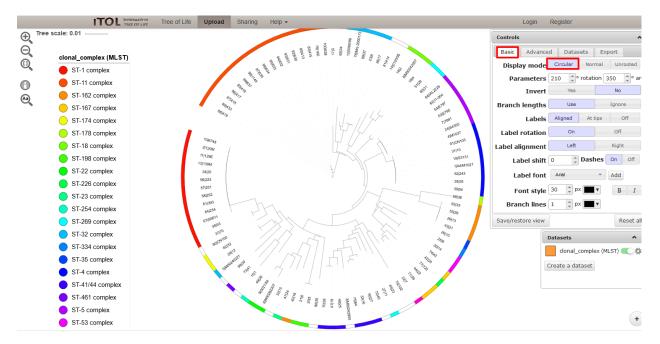
Your browser will open the iTOL website with your tree.



You can manipulate the tree in the browser, and display metadata by selecting the appropriate toggle.



The tree layout can be changed by clicking the 'Basic tab' and, for example, selecting a circular display mode.

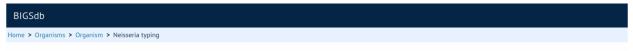


See the detailed documentation on the iTOL website for more information about manipulating and exporting trees.

14.10 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.

Expand the 'Analysis' section on the main contents page of a sequence definition database and click 'Locus Explorer'.



Neisseria typing database

Query a sequence	Find alleles	Search for allelic profiles	DI LOG IN	
Single sequence	By specific criteria	By specific criteria		
Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	L DOWNLOADS	+
matches.			EXPORT	+
Batch sequences	<u>By locus</u>	<u>By allelic profile</u>	MALYSIS	-
Query multiple independent sequences in FASTA format to identify allelic matches.	Select, analyse and download specific alleles from a single locus.	This can include partial matches to find related profiles.	Sequence similarity Sequence comparison	
		In a batch	Locus Explorer	
		Look up multiple allelic profiles together.		+
			③ INFORMATION	+
			SISOLATES	

14.10.1 Polymorphic site analysis

Select the locus you would like to analyse in the Locus dropdown box. The page will reload.

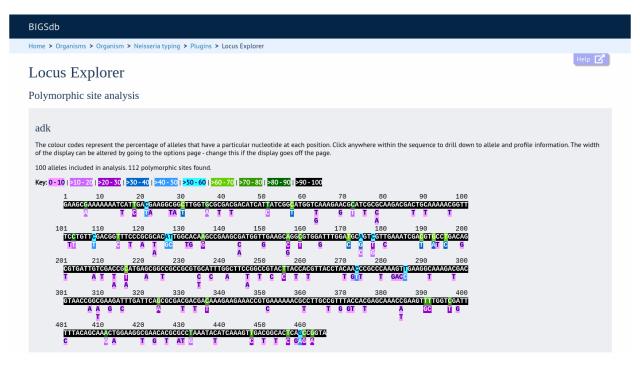
BIGSdb	
Home > Organisms > Organism > Neisseria typing > Plugins > Locus Explorer	
Locus Explorer	Help 🔽
Please select locus for analysis: Locus: adk Page will reload when changed Select sequencesSelect analysis	Action
Polymorphic Sites - Display polymorphic site frequencies and sequence schematic Codon - Calculate G+C content and codon usage Translate - Translate DNA to peptide sequences All None	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'.

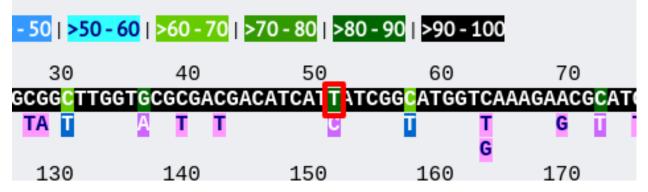
BIGSdb	
Home > Organisms > Organism > Neisseria typing > Plugins > Locus Explorer	
Locus Explorer	Help 🔀
Please select locus for analysis:	
Locus: adk Page will reload when changed	
Select sequences Select analysis	Action
98 99 99 100 101 102 102 103 All None	SUBMIT

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.



Along with the nucleotide frequencies, it will also show the percentage of allelic profiles containing each nucleotide at that position if the locus is part of a scheme such as MLST.

BIGSdb								
Home > Or	ganisms > Organ	ism > Neisseria	typing > Plugins > Loc	us Explorer				
Locu	s Explor	er						Help 🔽
Site I	Explorer							
	osition 51 es included in ana	lysis.						
Base Nu T	mber of alleles P 84	ercentage of alle 84.00	Les MLST profiles 14087 / 15577 (90.43%)					
С	16	16.00	349 / 15577 (2.24%)					

The second part of the page shows a table listing nucleotide frequencies at each of the variable positions.

Image: second definition of the system of the sy
Sittlon A C G T - %A %C %G %T %- 6 18 0 82 0 0 18.00 82.00 100 15 0 99 0 1 099.00 1.00 18 0 26 0 74.00 26.00 74.00
A \bullet C \bullet G \bullet T \bullet <t< th=""></t<>
15 0 99 0 1 0 99,00 1.00 18 0 26 0 74 0 26.00 74.00
18 0 26 0 74 0 26.00 74.00
21 0 59 0 41 0 59.00 41.00
22 3 0 97 0 0 3.00 97.00
27 0 99 0 1 0 99.00 1.00
28 1 0 99 0 0 1.00 99.00 <u>99.00</u>
30 0 62 0 38 0 62.00 38.00
36 18 0 82 0 0 18:00 82:00
39 0 99 0 1 0 99.00 1.00
42 0 98 0 2 0 98.00 2.00
51 0 16 0 84 0 16.00 84.00

See also:

- Investigating allele differences.
- Polymorphism analysis following isolate query.

14.10.2 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'.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Plugins > Locus Explorer	
Locus Explorer	Help 🔽
Please select locus for analysis:	
Locus: adk Page will reload when changed	
Select sequences Select analysis Action	
1 Polymorphic Sites - Display nolymorphic site frequencies and sequence schematic SUBMIT 2 © Codon - Calculate G+C content and codon usage SUBMIT 3 O Translate - Translate DNA to peptide sequences SUBMIT 4 5 0 6 •	
All None	

The GC content of the alleles will be determined and a table of the codon frequencies displayed.

BIGSdb					
lome > Orga	nisms > 0	rganism > Ne	isseria typing	> Plugins > L	cus Explorer
	Evel				Help 🔽
Locus	Expl	orer			
Codon	Usag	τe			
Bouon	obug	50			
adk					
ORF used: 1					
100 alleles i	included in	analysis.			
GC conter	nt				
Coding: GC 5					
1st letter: G					
2nd letter: G					
3rd letter: G	C 59.91%				
Codons					
Codons					
Codons Fraction: Pro	oportion of	usage of a giv	en codon amo	ong its redundar	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro	oportion of Jsage of gi	usage of a giv ven codon per	en codon amo 1000 codons.	ong its redundar	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U	Jsage of gi	ven codon per	1000 codons.		set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U Codon + A	Jsage of given mino acid	 Fraction + 	1000 codons. Frequency +	Number ¢	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Codon + A GCA	Jsage of gi [,] mino acid A	 Fraction ÷ 0.265 	1000 codons. Frequency + 17.484	Number + 271	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U Codon • A GCA GCC	Jsage of giv Amino acid A A	 Fraction + 0.265 0.235 	1000 codons. Frequency ¢ 17.484 15.484	Number + 271 240	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U Codon ¢ A GCA GCC GCG	Jsage of given and a second se	 Fraction + 0.265 0.235 0.399 	1000 codons. Frequency \$ 17.484 15.484 26.323	Number + 271 240 408	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Codon ¢ A GCA GCC GCG GCT	Jsage of given and a cide a ci	 Fraction + 0.265 0.235 0.399 0.101 	1000 codons. Frequency \$ 17.484 15.484 26.323 6.645	Number + 271 240 408 103	: set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U GCA GCA GCC GCG GCG GCT TGC	Jsage of given acid A A A A A A C	 Fraction \$ 0.265 0.235 0.399 0.101 0.970 	1000 codons. Frequency ⇒ 17.484 15.484 26.323 6.645 6.323	Number ♦ 271 240 408 103 98	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U Codon + A GCC GCC GCC GCC GCT TGT	Jsage of gir mino acid A A A A C C	 Fraction \$ 0.265 0.235 0.399 0.101 0.970 0.030 	1000 codons. Frequency ⇒ 17.484 15.484 26.323 6.645 6.323 0.194	Number + 271 240 408 103 98 3	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U Codon • A GCA GCC GCG GCT TGC TGC GAC	Jsage of given acid A A A A A A C C C D	 Fraction + 0.265 0.235 0.399 0.101 0.970 0.030 0.745 	1000 codons. Frequency \$ 17.484 15.484 26.323 6.645 6.323 0.194 91.097	Number + 271 240 408 103 98 3 1412	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Codon • A GCA GCC GCC GCC GCC TGC TGC TGT GAC GAC	Jsage of given acid A A A A A C C C C D D D	ven codon per Fraction ≠ 0.265 0.235 0.399 0.101 0.970 0.030 0.745 0.255	1000 codons. Frequency ≠ 17.484 15.484 26.323 6.645 6.323 0.194 91.097 31.226	Number ♦ 271 240 408 103 98 3 1412 484	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U Codon • A GCA GCG GCG GCG GCT TGT GAC GAA	Jsage of given acid A A A A C C C C D D D E	 ► Fraction ◆ ► Fraction ◆ ► 0.265 0.235 0.399 0.101 0.970 0.030 0.745 0.255 0.924 	1000 codons. Frequency \$ 17.484 15.484 26.323 6.645 6.323 0.194 91.097 31.226 83.355	Number ¢ 271 240 408 103 98 3 1412 484 1292	: set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Godon e A GCA GCC GCG GCG GCG GCG GCG GCA GCA GAA GA	Jsage of given acid A A A A C C C D D D D E E	 ► Fraction ► 0.265 0.235 0.399 0.101 0.970 0.030 0.745 0.255 0.924 0.076 	1000 codons. Frequency ¢ 17.484 15.484 26.323 6.645 6.323 0.194 91.097 31.226 83.355 6.839	Number 271 240 408 103 98 3 1412 484 1292 106	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: D GCA GCC GCC GCC GCC GCC TGC TGC TGC TGC GAA GAA GAA GAA GAA GAA	Jsage of given acid A A A A C C C D D D E E E E F	 ► Fraction ► 0.265 0.235 0.399 0.101 0.970 0.030 0.745 0.255 0.924 0.076 0.614 	1000 codons. Frequency ↓ 17.484 15.484 26.323 6.645 6.323 0.194 91.097 31.226 83.355 6.839 15.806	Number ◆ 271 240 408 103 98 3 1412 484 1292 106 245	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Prc Frequency: L GCA GCA GCC GCC GCC GCT TGC GAC GAC GAA GAA GAG TTC TTT	Jsage of gi mino acid A A A A A A C C C D D D E E E E F F	 ► Fraction ► 0.265 0.235 0.399 0.101 0.970 0.030 0.745 0.255 0.924 0.076 0.614 0.386 	1000 codons. Frequency ≠ 17.484 15.484 26.523 6.645 6.323 0.194 91.097 31.226 83.355 6.839 15.806 9.935	Number ◆ 271 240 408 103 98 3 1412 484 1292 106 245 154	set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: D GCA GCC GCC GCC GCC GCC TGC TGC TGC TGC GAA GAA GAA GAA GAA GAA	Jsage of given acid A A A A C C C D D D E E E E F	 ► Fraction ► 0.265 0.235 0.399 0.101 0.970 0.030 0.745 0.255 0.924 0.076 0.614 	1000 codons. Frequency ↓ 17.484 15.484 26.323 6.645 6.323 0.194 91.097 31.226 83.355 6.839 15.806	Number ◆ 271 240 408 103 98 3 1412 484 1292 106 245	set (i.e. the set of codons which code for this codon's amino acid).

14.10.3 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'.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Plugins > Locus Explorer	
Locus Explorer	Help 🔽
Please select locus for analysis:	
Locus: adk Page will reload when changed	
Select sequences Select analysis Action	
1 • Polymorphic Sites - Display polymorphic site frequencies and sequence schematic SUBMIT 2 • Translate - Translate DNA to peptide sequences SUBMIT 3 • Translate - Translate DNA to peptide sequences SUBMIT 4 • Translate - Translate DNA to peptide sequences • Translate DNA to peptide sequences All None None • Translate DNA to peptide sequences	

An aligned amino acid sequence will be displayed.

BIGSdb									
Home > Organis	sms > Organism >	Neisseria typin	g > Plugins	> Locus Explo	rer				
Locus E	Explorer								Help 🗹
Translat	te - aligne	ed prote	ein sea	liences					
Tunsiat	ungin	Lu prou	in seq	uciicei	,				
adk ORF used: 1									
The width of th	he alignment can be	e varied by goin	g to the optio	ns page.					
100 alleles incl	luded in analysis.								
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	10 	20						······	

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.

14.11 Microreact

Microreact is a tool for visualising genomic epidemiology and phylogeography. Individual nodes on a displayed tree are linked to nodes on a geographical map and/or timeline, making any spatial and temporal relationships between isolates apparent.

The Microreact plugin generates Neighbour-joining trees from concatenated sequences for any selection of loci or schemes and uploads these, together with country and year field values to the Microreact website for display.

Note: While Microreact itself is able to display isolates using GPS coordinates, the BIGSdb plugin is currently limited to the level of country.

Microreact can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria isolates		
Neisseria isolates database		
Query database	HO LOG IN	
Search database		
Browse, search by any criteria, or enter list of attributes.	PROJECTS	+
	EXPORT	+
Search by combinations of loci This can include partial matches to find related isolates.	🗠 ANALYSIS	
		+
		+
	S TYPING	

Jump to the 'Third party' category, follow the link to BLAST, then click 'Launch Microreact'.

Summary: Open data visualization and sharing for genomic epidemiology Microreact is a tool for visualising genomic epidemiology and phylogeography (Argimón <i>et al</i> 2016 <i>Microb Genom</i> 2:e000093). Individual nodes on a displayed tree are linked to nodes on a geographical map and/or timeline, making any spatia and temporal relationships between isolates apparent. The Microreact Julyin generates Neighbour-joining trees from concatenated sequences for any selection of loci or schemes and uploads these, together with country and year field values, to the Microreact website for display. Documentation Digdbreadthedocsio C
geographical map and/or timeline, making any spatial and temporal relationships between isolates apparent. The Microreact plugin generates Neighbour-joining trees from concatenated sequences for any selection of loci or schemes and uploads these, together with country and year field values, to the Microreact website for display. Documentation bigsbbreadhedocsio C

Alternatively, it can be accessed following a query by clicking the 'Microreact' button at the bottom of the results table. Isolates returned from the query will be automatically selected within the Microreact plugin interface.

35	26	B278; NIBSC_2764; Z1278	Niger	1963	invasive (unspecified/other)	Neisseria meningitidis	А	4	ST-4 complex	7	13	F1-5	
52	243	B362; NIBSC_2779; Z1362	Cameroon	1966		Neisseria meningitidis	A	4	ST-4 complex	7	13	F1-5	
61	393	B392; NIBSC_2823; Z1392	Greece	1968	carrier	Neisseria meningitidis	А	1	ST-1 complex	5-2	10	F5-1	
64	254	B439; NIBSC_2812; Z1439	Djibouti	1966	invasive (unspecified/other)	Neisseria meningitidis	Α	1	ST-1 complex	5-2	10	F1-7	
67	S5611	B466; NIBSC_2765; Z1466	Australia	1977	invasive (unspecified/other)	Neisseria meningitidis	Α	1	ST-1 complex	5-2	10	F5-1	
82	11-004	B503; NIBSC_2826; Z1503	China	1984	invasive (unspecified/other)	Neisseria meningitidis	Α	5	ST-5 complex	20	9	F3-8	
84	IAL2229	B506; NIBSC_2816; Z1506	Brazil	1976		Neisseria meningitidis	Α	5	ST-5 complex	20	9	F2-1	
90	CN100	B534; NIBSC_2768; Z1534	UK [England]	1941	invasive (unspecified/other)	Neisseria meningitidis	A	21		5-2	10	F3-9	
120	F4698	NIBSC_2731; Z3515	Saudi Arabia	1987	carrier	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1	
128	F6124	NIBSC_2730; Z3524	Chad	1988	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1	
160	1014	NIBSC_2821; Z3667	Sudan	1985	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5	
210	H1964	NIBSC_2796; Z3771	UK	1987	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1	
237	H44/76	44/76-3; NIBSC_2724; Z3842	Norway	1976	invasive (unspecified/other)	Neisseria meningitidis	В	32	ST-32 complex	7	16	F3-3	
238	153	NIBSC_2733; Z3905	China	1966	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1	
«		1 2 🕡 🔊											
Δr	nalysis	tools											

Rreakdown:	Fields	Two Field	Combinations	Polymorphic sites	Publication	s Sequence bin	
🛃 Analysis:	BURST	Codons	Gene Presence	Genome Comparator	BLAST	rMLST species id	PCR
Export:	Dataset	Contigs	Sequences				
🗹 Third party:	GrapeTr	ee iTOL	PhyloViz Mic	roreact			

Select the isolates to include. The tree can be generated from allelic profiles of any selection of loci, or more conveniently, you can select a scheme in the scheme selector, or from a list of recommended schemes if these have been set, to include all loci belonging to that scheme.

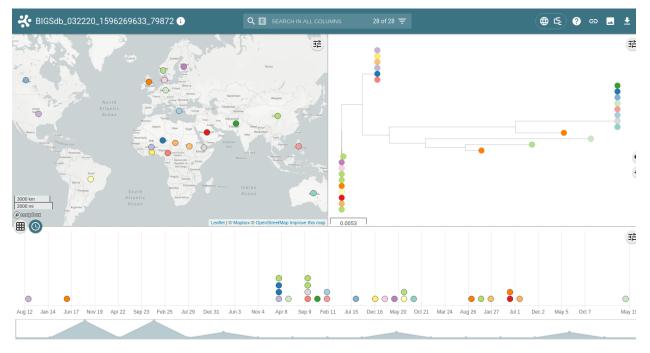
Additional fields can be selected to be included as metadata for use in colouring nodes - select any fields you wish to include. Multiple selections can be made by holding down shift or ctrl while selecting. Click 'Submit' to start the analysis.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Microreact		
Microreact - Neisseria isolates		Help 🗹 🚯
This tool will generate neighbor-joining trees from concatenated nucleotide sequences peptide loci with genome sequences, can be included. Please check the loci that you wo	epidemiology and phylogeography. <i>Micro</i> Only DNA loci that have a correspondin	g database containing allele sequence identifiers, or DNA and
scheme. Analysis is limited to 3,000 records or 120,000 sequences (records x loci). Isolates Loci 12 13 10 11 13 10 Clear List all All None Paste list	Recommended schemes Select one or more schemes below or use the full schemes list. MLST Ribosomal MLST N. meningitidis cgMLST v1.0 N. gonorrhoeae cgMLST v1.0	Schemes
Descriptions	Include fields	Action
Modify the values below - these will be displayed within the created Microreact proje Title: Description: Neisseria isolates	ct. Select additional fields to include (isolate, country and year are alwa continent region date sampled isoyear sampled week sampled date received	

The job will be sent to the job queue. When it has finished, click the button marked 'Launch Microreact'.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Job status viewer
Job status viewer
Status BiGSdb_032220_1596269633_79872 Submit time: 2020-08-01 08:13:53 Status Finished Start time: 2020-08-01 08:14:00 Progress: 100% Stop time: 2020-08-01 08:14:08 Total time: 8 seconds
Output Launch Microreact Files
Concatenated FASTA
Tar file containing all output files
Please note that job results will remain on the server for 7 days.

The generated tree will be uploaded to the Microreact website and displayed. Clicking any node will show its position(s) within the tree, map and timeline. A node on the map may correspond to multiple nodes in the tree or timeline.



14.12 PhyloViz

PhyloViz Online is a tool for generating and visualising minimum-spanning trees. Datasets can include metadata which allows nodes in the resultant tree to be coloured interactively.

PhyloViz can be accessed by selecting the 'Analysis' section on the main contents page.

BIG	SSdb		
Hom	ne > Organisms > Neisseria spp. > Neisseria isolates		
N	<i>leisseria</i> isolates database		
	Query database	D LOG IN	
	Search database		
	Browse, search by any criteria, or enter list of attributes.	PROJECTS	+
		Z EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.	MALYSIS	
			+
		O INFORMATION	+

Jump to the 'Third party' category, follow the link to PhyloViz, then click 'Launch PhyloViz'.

PhyloViz Summary: Creates phylogenetic inference and data visualization for sequence-based typing methods PhyloViz Online is a tool for generating and visualising minimum-spanning trees based on allelic profiles. This plugin generates datasets that are then uploaded to PhyloViz Online for visualisation. Commentation Specificationations of the resultant tree to be coloured interactively. Locurentation Specificationations Locure PhyloViz Locure PhyloViz

Alternatively, it can be accessed following a query by clicking the 'PhyloViz' button at the bottom of the results table. Isolates returned from the query will be automatically selected within the PhyloViz interface.

15	1			Germany	1999	carrier		Neisseria meningi	tidis	E	864				
16	2			Germany	1999	carrier	r	Neisseria meningi	tidis	В	854	ST-18 complex			
17	3			Germany	1999	carrier	r	Neisseria meningi	tidis	W	174	ST-174 complex			
18	4			Germany	1999	carrier	1	Neisseria meningi	tidis	В	19	ST-18 complex			
19	S3131	B213; NIBSC_28	313; Z1213	Ghana	1973 inv	vasive (unspeci	ified/other)	Neisseria meningi	tidis	A	4	ST-4 complex	7	13-1	F1-5
20	5			Germany	1999	carrier		Neisseria meningi	tidis	NG	198	ST-198 complex			
21	6			Germany	1999	carrier	r	Neisseria meningi	tidis	NG	198	ST-198 complex			
22	7			Germany	1999	carrier	1	Neisseria meningi	tidis	E	60	ST-60 complex			
23	8			Germany	1999	carrier		Neisseria meningi	tidis	В	32	ST-32 complex			
	C 4755	B227; NIBSC 28	306.71227	Denmark	1974 inv	vasive (unspeci	ified/other)	Neisseria meningi	tidis	Α	5	ST-5 complex	5-1	9	F3-1
24	S4355	DZZ/, NIDSC_ZC	500, 21227												
	9	BZZ7, NIBSC_20	500, 21227	Germany	1999	carrier		Neisseria meningi	tidis	В	930	ST-334 complex			
25 («		1 2 3	4 5 6					2	tidis	В	930	ST-334 complex			
25 « An	9) (1 2 3 cools	4 5 6		1999	carrier		Neisseria meningi	tidis	В	930	ST-334 complex			

Export: Dataset Contigs Sequences

Third party: GrapeTree iTOL PhyloViz Microreact

Select the isolates to include. The tree can be generated from allelic profiles of any selection of loci, or more conveniently, you can select a scheme in the scheme selector to include all loci belong to that scheme.

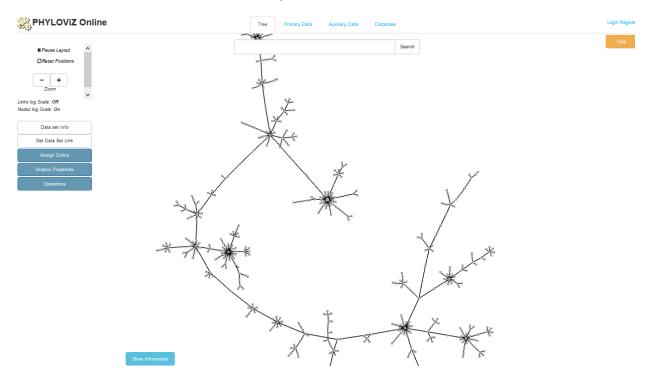
Provenance fields can be selected to be included as metadata for use in colouring nodes - select any fields you wish to include. Multiple selections can be made by holding down Shift or Ctrl while selecting. Click 'Submit' to start the analysis.

BIGSdb				
Home > Organisms > Neisseria spp. > Neisseria iso	olates > Plugins > PhyloViz			
PhyloViz: phylogenetic	tree vizualisatio	on - Neisseria isolates		Help 🕜 🚯
PHYLOVIZ Online is developed 1. Instituto de Microbiologi 2. INESC-ID, R. Alves Redol § 3. Instituto Superior Técnice 4. Instituto Superior de Eng Web site: https://online.phyloviz.net/	a, Instituto de Medicina Molecula 9, 1000-029 Lisboa, Portugal 1, Universidade de Lisboa, Lisboa, enharia de Lisboa, Instituto Polité	dré Carriço (1), Alexandre P. Francisco (2,3), Cátia Vaz r, Faculdade de Medicina, Universidade de Lisboa, Li Portugal écnico de Lisboa, 007 Lisboa, Portugal	sboa, Portugal	
Optionally select provenance fields to include in the Select ids	ne dataset. This allows you to colo	our nodes based on field values. Analysis is limited t	o 10,000 isolates.	
Action	date sampled isoyear sampled isoyear sampled	IGS_TDNA IGS_TRNA (SSU_rRNA) 23S_TRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) All None Paste list	A Retabulstit A Retabulstit A Retabulstit A Reserved AMR A Reserved	* ,

The necessary files will be generated immediately. When this has finished, click the button launch 'Launch PhyloViz'.



The tree will be sent to and rendered within the PhyloViz website.



See more information about manipulating the tree on the PhyloViz website.

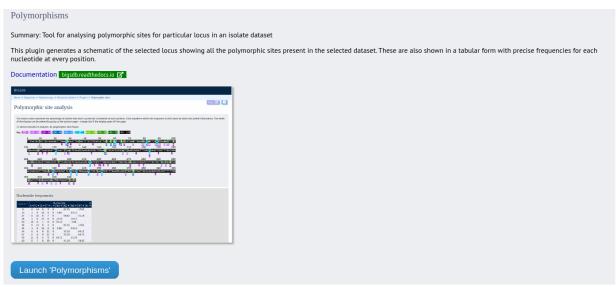
14.13 Polymorphisms

The Polymorphisms plugin generates a *Locus Explorer* polymorphic site analysis on the alleles designated in an isolate dataset following a query.

The analysis can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb			
Home > Organisms	; > Neisseria spp. > Neisseria isolates		
Neisseria	a isolates database		
	Query database	→) LOG IN	
	Search database		
	Browse, search by any criteria, or enter list of attributes.	PROJECTS	+
		EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.	🗠 ANALYSIS	
			+
		1 INFORMATION	+
		S TYPING	

Jump to the 'Breakdown' category, follow the link to Polymorphisms, then click 'Launch Polymorphisms'.



The analysis is accessed by clicking the 'Polymorphic sites' button in the Breakdown list at the bottom of a results table following a query.

	-		Germany	1999	carrier	Neissena meningitiois	E	804	CT 40			
6	2		Germany	1999	carrier	Neisseria meningitidis	В	854	ST-18 complex			
7	3		Germany	1999	carrier	Neisseria meningitidis	W		ST-174 complex			
.8	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
.9		B213; NIBSC_2813; Z121			wasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
0	5		Germany	1999	carrier	Neisseria meningitidis	NG		ST-198 complex			
1	6		Germany	1999	carrier	Neisseria meningitidis	NG		ST-198 complex			
2	7		Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningitidis	В	32	ST-32 complex			
.4		B227; NIBSC_2806; Z1222			vasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	5-1	9	F3-1
5	9		Germany	1999	carrier	Neisseria meningitidis	В	930	ST-334 complex			
@	0	1234 (>									
@) () nalysis t		00									
(« Ar) () nalysis t	ools	00									
(« Ar) () nalysis t		Combinations		orphic sites Publication							
(« Ar) () nalysis t	OOlS wn: Fields Two Field	00	Polyme	orphic sites Publication							
(« Ar) () nalysis t () Breakdo	OOlS wn: Fields Two Field ysis: BURST Codons	Combinations	Polyme		s Sequence bin						

Select the locus that you would like to analyse from the list.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria is	olates > Plugins > Polymorphic sites	
Polymorphic site analys	is	Help 🔀 🚺
		w minutes (or longer depending on the queue). This is because sequences n.
Loci	Options	Action
165_rDNA 165 rRNA (SSU rRNA) 235_rRNA abcZ abcZ (NEIS1015) accF2 (NEIS1279) acck22 (NEIS1727) ack422 (NEIS1727)	If both allele designations and tagged sequences exist for a locus, choose how you want these handled: ① © Use allele sequence retrieved from external database ○ Use sequences tagged from the bin © Analyse single example of each unique sequence	ANALYSE
acnA (NEIS1729)	Exclude incomplete sequences	

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.

olymo																		Help 🔽	71
	orp	hic	sit	e a	ana	lysi	S												
The colour co of the displa													/here within tl	ne sequence	to drill dow	n to allele	and profile i	information. The	wid
17 alleles ind	cludeo	in ana	lysis.	31 po	lymorp	hic site	es found	d.											
Key: <mark>0 - 10</mark>	>10 - 2	0 >20	- 30	>30 -	40 >4	0 - 50	>50-6	<mark>50</mark> >60 - 7	0 >70-8	>80 - 9	0 >90-10	00							
1		10		20	0	3	0	40		50	60	70	80	90	100				
	TGAT		TGCC		GTTTG	GG <mark>C</mark> GA	AATTC		ATTGCGCO		A <mark>tcat</mark> gto	AGCCATGAGT1	GGAAAA <mark>C</mark> GG	TTCG <mark>G</mark> GTGA					
101		110		120	<mark>۵</mark>	13		140	A 15	<mark>2</mark>	C GC 1	G I I 170	GC TAA 180	190					
	AAAG		AACG									GCAGTCAAGCAG							
A	N.	G	Т		G	CCG			T	AT		G G A	G	I	A				
201		210		220	0	23	0	240	25	i0	260	270	280	290	300				
	AA <mark>C</mark> GA		TCGG						G <mark>T</mark> GTCGC0	TTGGCG	CAGGCTTO	GGTGCA <mark>G</mark> AAGC	CCGACGTAT		ACGA <mark>A</mark> CCG				
301	Ū.	310		320	ຼີ	33	A	340	G T 35	CA T	360	370	T T 380	390	G 400				
	CAAC		GGAT									AGCTTGGTTGT							
			С		Т		C 0		i i	G		C C		Т	C				
401		410		420	0	43	0												
AC	AATA	rece																	
				-	TGTCG														
Ū		T Ū		-	TGTCG C T	AACTC T													
ū				-	_														
	1	Τü	A	C	_														
T Nucleoti	1	Τü	A	C	_														
	1	Τü	A	es	3 T	Ū													
Nucleoti Position +	ide f	T I	A enci	es Nu	C T	te	C	0/T • 0/	•										
Nucleoti Position +	ide f	T T requ	A enci	es Nu - +	3 T	de %C ÷		%T + %-	* *										
Nucleoti Position +	ide f	T u requ	A enci	es Nu	C T	te	C %G ÷	%T ♦ %- 17.65	¢ \$										
Nucleoti Position + 9	ide f	T u requ	A enci	C (es • ↓ 0	ucleotic	de %C ÷	C		¢ ¢										
Nucleoti	ide f A + 0 0 1 0 4	T U requ 4 0 0 16 0 0 13	A enci 5 T ÷ 3 0 7 0	C (es - + 0 0 0 0	C T ucleotic %A ↔ 5.88 23.53	le %C \$ 82.35	€ %G ◆ 94.12 76.47	17.65	¢ •										
Nucleoti Position + 9 24 27 28 30	ide f 0 1 0 4 16	t e q u u u u u u u u	A enci 3 0 7 0 0	C (es - +	ucleotic %A ÷ 5.88	te %C ¢ 82.35 58.82	€ %G ≑ 94.12	17.65 41.18	•										
Nucleoti Position + 9 24 27 28 30 36	ide f 0 1 0 4 16 0	T G G G G G G G G G G G G G G G G G G G	A enci 3 0 7 0 0 3	C (es - + 0 0 0 0 0 0 0 0	C T ucleotic %A + 5.88 23.53 94.12	le %C \$ 82.35	%G ↓ 94.12 76.47 5.88	17.65	•										
Nucleoti 9 24 27 28 30 36 45	ide f 0 1 0 4 16 0 1	t e q u u u u u u u u	A enci 3 0 7 0 0 3 0 3 0	C (es - +	C T ucleotic %A ↔ 5.88 23.53	te %C \$ 82.35 58.82 82.35	€ %G ◆ 94.12 76.47	17.65 41.18	•										
Nucleoti Position + 9 24 27 28 30 36	ide f 0 1 0 4 16 0 1 0	T G G G G G G G G G G G G G G G G G G G	A Enci 3 0 7 0 0 3 0 3 0	C (C (C (C (C (C (C (C (C T ucleotic %A + 5.88 23.53 94.12	te %C ¢ 82.35 58.82	%G ↓ 94.12 76.47 5.88	17.65 41.18 17.65											
Nucleoti Position • 9 24 27 28 30 36 45 54 57 59	ide f A + 0 0 1 0 4 16 0 1 0 1 0 1 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1	T I requ + G .4 0 0 13 0 13 0 14 0 16 0 16 0 16 0 16 0 16 0 16 0 16 0 16 0 16 0 16 0 16 0 16 0 16 0 16	A T + 3 0 7 0 0 3 0 11 111 0	E S • • • • 0	C T ucleotic %A + 5.88 23.53 94.12	de %C + 82.35 58.82 82.35 35.29 35.29	%G ↓ 94.12 76.47 5.88	17.65 41.18 17.65 64.71 64.71											
Nucleoti Position • 9 24 27 28 30 36 45 54 57 59 60	ide f A + 0 0 1 1 0 4 1 1 0 0 1 0 0 1 0 0 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0	T I + G & +	A 2 2 2 2 2 4 2 4 2 4 2 4 4 4 4 4 4 4 4 4 4 4 4 4	€ S Nu - ↓ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	C T UCLEOTIO %A + 5.88 23.53 94.12 5.88	te %C + 82.35 58.82 82.35 35.29 35.29 41.18	%G ◆ 94.12 76.47 5.88 94.12	17.65 41.18 17.65 64.71 64.71 58.82	•										
Nucleoti Position + 9 24 27 28 30 36 45 54 57 59 60 63	ide f A	T I + G 4 .4 0 .4 0 .0 13 .0 14 .4 0 .6 0 .6 0 .6 0 .6 0	A 2 2 2 2 2 4 2 4 2 4 2 4 4 4 4 4 4 4 4 4 4 4 4 4	€ S Nu - ↓ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	 Cleotic %A + 5.88 23.53 94.12 5.88 64.71 	de %C + 82.35 58.82 82.35 35.29 35.29	 %G ◆ 94.12 76.47 5.88 94.12 35.29 	17.65 41.18 17.65 64.71 64.71	•										
Nucleoti Position • 9 24 27 28 30 36 45 54 57 59 60 63 64	ide f A	T I ↓ G € ↓ 4 00 14 00 16 ↓ 0	A enci T + 3 0 7 0 0 0 111 111 0 10 10	es Nu - + 0 0 0 0 0 0 0 0 0 0 0 0 0	 C T S.88 S.88 S.88 G4.71 G4.71 	le %C ↓ 82.35 58.82 82.35 558.82 82.35 35.29 35.29 41.18 94.12	%G ◆ 94.12 76.47 5.88 94.12	17.65 41.18 17.65 64.71 64.71 58.82 5.88											
Nucleoti Position + 9 24 27 28 30 36 45 54 57 59 60 63	ide f A + 0 0 1 0 4 16 0 1 0 0 11 0 0 11 0 0 11 0 0 11 0 0 0 1 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0	T I + G 4 .4 0 .4 0 .0 13 .0 14 .4 0 .6 0 .6 0 .6 0 .6 0	A enci T ¢ 3 0 7 0 0 111 111 0 10 1 0 4	€ S Nu - ↓ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	 C T S.88 S.88 S.88 G4.71 G4.71 	te %C + 82.35 58.82 82.35 35.29 35.29 41.18	 %G ◆ 94.12 76.47 5.88 94.12 35.29 	17.65 41.18 17.65 64.71 64.71 58.82 5.88 23.53											
Nucleoti Position • 9 24 27 28 30 36 45 54 57 59 60 63 64 66 64 66 72	ide f 0 1 0 4 16 0 11 0 0 11 0 0 11 0 0 0 11 0 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 0 0 1 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0	T I ↓ G € ↓ 0 <t< td=""><td>A 2 T ¢ 3 0 7 0 0 0 3 0 111 110 10 10 4 3</td><td>C C</td><td> C T S.88 S.88 S.88 G4.71 G4.71 </td><td>le %C ↓ 82.35 58.82 82.35 558.82 82.35 35.29 35.29 41.18 94.12</td><td>%G ◆ 94.12 76.47 5.88 94.12 35.29</td><td>17.65 41.18 17.65 64.71 64.71 58.82 5.88 23.53</td><td>•</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	A 2 T ¢ 3 0 7 0 0 0 3 0 111 110 10 10 4 3	C C	 C T S.88 S.88 S.88 G4.71 G4.71 	le %C ↓ 82.35 58.82 82.35 558.82 82.35 35.29 35.29 41.18 94.12	%G ◆ 94.12 76.47 5.88 94.12 35.29	17.65 41.18 17.65 64.71 64.71 58.82 5.88 23.53	•										
Nucleoti Position • 9 24 27 28 30 36 45 54 57 59 60 63 64 66 72 78	A + 0 0 1 0 1 0 1 0 1 0 11 0 0 11 0 0 11 0 0 11 0 0 11 0 0 11 0 111 0 0 114	Image: Total and the second	A T + 3 0 7 0 0 3 0 111 110 100 100 100 4 4 3 0	C C	 T xA + 5.88 5.88 64.71 64.71 	le %C + 82.35 58.82 82.35 35.29 41.18 94.12 76.47	%G + 94.12 76.47 5.88 94.12 35.29 35.29 82.35	17.65 41.18 17.65 64.71 64.71 58.82 5.88 23.53											

14.14 Species identification

The species identification tool extracts ribosomal MLST alleles from genomes and determines the species based on the count of alleles that are uniquely found within a single species (or higher taxonomic rank if alleles unique to a species are not found). This is done by making a query to the rMLST genome database.

The tool can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb					
Home > Organi	ms > Neisseria spp. > Neisseria isolates				
Neisse	<i>ia</i> isolates database				
	Query database	DILOG IN			
	Search database				
	Browse, search by any criteria, or enter list of attributes.	PROJECTS	+		
		Z EXPORT	+		
	Search by combinations of loci This can include partial matches to find related isolates.	MALYSIS			
			+		
			+		

Jump to the 'Analysis' category, follow the link to rMLST species identity, then click 'Launch rMLST species identity'.

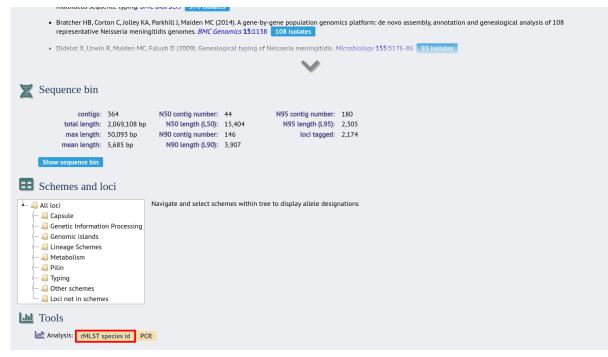
The tool can be accessed from the front page of an isolate database.

rMLST species identity
Summary: Use rMLST to identify species from bacterial genome assemblies
The species identification tool extracts ribosomal MLST alleles from genomes and determines the species based on the count of alleles that are uniquely found within a single species (or higher taxonomic rank if alleles unique to a species are not found). This is done by making a query via the PubMLST RESTful API to the rMLST typing database.
Documentation bigsdbreadthedocs io 🗗

Alternatively, it can be accessed following a query by clicking the 'rMLST species id' button at the bottom of the results table. Isolates returned from the query will be automatically selected within the species id interface (note that only isolates with a genome assembly will be able to be checked).

	Display,	/sort options —					Seque Include old recor Add filter: Action		uen	ce bin size >= 2	Mbp 🖌	€ ✓ Add	
		r by: id olay: 25 🗸 reco	ırds per pa	ige 🚺		✓ ascending	• RESET	SEARCH					
7	ecords retu	irned. Click the hy	/perlinks fo	or detailed i									
						te fields 👔				MLST		typing anti	
d 1	isolate A4/M1027	aliases B1; NIBSC 280		country USA	year	disease invasive (unspecified/other)	species	capsule group A	SI	clonal complex	PorA_VR 5-2	PorA_VR2	FetA_VR F1-5
2	120M	B35; NIBSC 282		Pakistan		meningitis and septicaemia		A	1	ST-1 complex	5-2	10	F5-1
7	7891	B54; NIBSC 276		Finland		invasive (unspecified/other)		A		ST-5 complex	20	9	F3-1
0	6748	B73; NIBSC_278		Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	18-1	3	F5-1
1	129E	B92; NIBSC_282	8; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	А	1	ST-1 complex	5-2	10	F3-6
3	139M	B99; NIBSC_279					Neisseria meningitidis	A		ST-1 complex	5-2	10	F5-1
9	S3131	B213; NIBSC_28	13; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
A	nalysis () Breakd		Two Field	Combina	tions	Polymorphic sites Public	ations Sequence bin						
	1.0												
	🛃 Ana	alysis: BURST	Codons	Gene Pres	ence	Genome Comparator BLA	ST rMLST species id	PCR					
	🛃 Ana	-	Codons Contigs			Genome Comparator BLA	ST rMLST species id	PCR					

Finally, the analysis is also possible directly from an isolate record, if the isolate has a genome assembly associated with it.



The tool interface consists of a list of isolate ids to check. This will be pre-populated if accessed following a query or directly from an isolate record. If the rMLST scheme is defined on the system, you will have a choice as to whether to BLAST the genome sequences to identify the rMLST alleles, or just use the designations that are tagged in the database. The latter is much quicker but relies on the record having been scanned and annotated with the rMLST loci.

Click 'Submit'.

BIGSdb			
Home > Organisms > Neisseria sp	p. > Neisseria isolates > Plugins > Spe	ies identification	
rMLST species	identification		
validated genomes in the rMLS	T database.	ected isolate sequence record(s). A predicted taxon will be s These isolate records must include genome sequences.	hown where identified alleles have been linked to
Isolates	Options	Action	
1 2 7 10 11 13 19 Clear List all	 Scan genome or use allele designations already stored Scan genomes Use stored allele designations 	SUBMIT	

The job will be sent to the job queue.

Results will be displayed in a table as they are generated. The table will display the highest taxonomic rank that can be reliably identified, e.g. species, the taxon and its full taxonomy. An indication of the confidence for the result will also be displayed - this is based on the proportion of alleles found that are unique to a taxon.

	11-4-			Prediction from identified rMLST alleles linked to genomes			Identified rSTs
	isolate	Rank	Taxon	Taxonomy	Support Matches	rST	Species
1	A4/M1027	SPECIES	Neisseria meningitidis	Proteobacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria meningitidis	100%	2504	Neisseria meningitidis
2	120M	SPECIES	Neisseria meningitidis	Proteobacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria meningitidis	100%	2532	Neisseria meningitidis
7	7891	SPECIES	Neisseria meningitidis	Proteobacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria meningitidis	100%	2434	Neisseria meningitidis
10	6748	SPECIES	Neisseria meningitidis	Proteobacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria meningitidis	100%	2530	Neisseria meningitidis
11	129E	SPECIES	Neisseria meningitidis	Proteobacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria meningitidis	100%	2528	Neisseria meningitidis
13	139M	SPECIES	Neisseria meningitidis	Proteobacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria meningitidis	100%	2433	Neisseria meningitidis
19	S3131	SPECIES	Neisseria meningitidis	Proteobacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria meningitidis	100%	2285	Neisseria meningitidis
ile		eport file (JS	ON format)				

Clicking the '+' icon on any row will display further details about the matches.

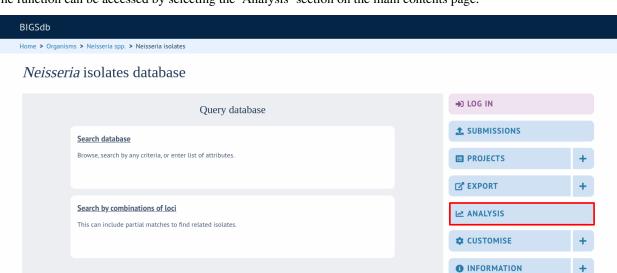
	Inclote						Prediction from identified rMLST alleles linked to genomes			dentified rSTs			
id	isolate	Rank	(Taxor	า		TaxonomySuppo	rt Matches	rST	Species			
1	A4/M1027	SPECI		Neisse. meningi		Proteoba meningit	tteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria dis	• •	2504	Neisseria meningitidis			
2	120M	SPECI		Neisse. meningi		Proteobal	obacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria ingitidis						
7	7891	SPECI	SPECIES Neisseria meningitidis SPECIES Neisseria			Proteobal meningit	eobacteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria 100%						
10	6748	SPECI		Neisse meningi		Proteobal meningit	rteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria dis	•	2530	Neisseria meningitidis			
53	loci matche	ed (rML)	ST uses	53 in tot	al)								
	Locus	Allele	Length	Contig	Start position	End position	Linked data values						
ΒA	CT000001	174	1680	181603	32429	34108	species: Neisseria meningitidis [n=1830]						
BA	CT000002	9	729	181098	6971	7699	species: Neisseria meningitidis [n=694]						
BA	CT000003	1	693	181315	13840	14532	species: Neisseria meningitidis [n=1772]						
BA	CT000003	1	693	181474	5814	6506	species: Neisseria meningitidis [n=1772]						
BA	CT000004	2	621	181315	5569	6189	species: Neisseria meningitidis [n=7822]; Neisseria sp. [n=1]						
BA	CT000005	1	519	181315	9345	9863	species: Neisseria meningitidis [n=17221]; Neisseria gonorrhoeae [n=1]; Neisseria sp. [n=1]						
	CT000005	1		181474		11001	species: Neisseria meningitidis [n=17221]; Neisseria gonorrhoeae [n=1]; Neisseria sp. [n=1]						
	CT000006	5		181603			species: Neisseria meningitidis [n=2053]; Neisseria sp. [n=1]						
	CT000007	1		181145	2253	2723	species: Neisseria meningitidis [n=14753]; Neisseria sp. [n=1]						
	CT000007	1		181374	2271	2741	species: Neisseria meningitidis [n=14753]; Neisseria sp. [n=1]						
	CT000008	2		181315	10796	11188	species: Neisseria meningitidis [n=7491]						
	CT000009	1		181193		17327	species: Neisseria meningitidis [n=3125]						
	CT000009	1	- / -	181591			species: Neisseria meningitidis [n=3125]						
	CT000010	1		181315			species: Neisseria meningitidis [n=926]						
	CT000010	1		181474	202	513	species: Neisseria meningitidis [n=926]						
	CT000011	2		181315	6209	6604	species: Neisseria gonorrhoeae [n=15852]; Neisseria meningitidis [n=7271]; Neisseria sp. [n=1]						
	CT000012	1		181145	2841	3212	species: Neisseria meningitidis [n=1337]; Neisseria gonorrhoeae [n=3]						
_	CT000012	1		181374	2859	3230	species: Neisseria meningitidis [n=1337]; Neisseria gonorrhoeae [n=3]						
BA	CT000013	1	363	181315	6624	6986	species: Neisseria meningitidis [n=19291]; Neisseria sp. [n=1]						

Note: Ribosomal MLST was first described in Jolley et al. 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 158:1005-15

14.15 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 accessed by selecting the 'Analysis' section on the main contents page.



Jump to the 'Breakdown' category, follow the link to Sequence Bin Breakdown, then click 'Launch Sequence Bin Breakdown'.

The function can be selected by clicking the 'Sequence bin' link on the Breakdown section of the main contents page.

quence Bin Breakdown	
mmary: Breakdown of sequence bin contig properties	
e 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 each isolate record. Using this la tric against a cgMLST scheme can be a good indicator of genome quality. Values for number of contigs, total sequence length, mean contig length and contig length distribution are charted.	tter
cumentation bigsdb.readthedocs.io 😰	
Stel > topsets + Materials > Materials > Materials to Materials Totalisment eakdown of sequence bin contig properties	
Impute the second sec	
A PARAMETARY A PAR	
Launch 'Sequence Bin Breakdown'	

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.

19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
24	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	5-1	9	F3-1
30	14	BennettTree10; alpha14	Germany	1999	carrier	Neisseria meningitidis	cnl	53	ST-53 complex	7	30-3	F5-5
31	10	B269; NIBSC_2825; Z1269	Burkina Faso	1963	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
34	20	B275; NIBSC_2767; Z1275	Niger	1963	invasive (unspecified/other)	Neisseria meningitidis	A			5-2	10	F1-7
35	26	B278; NIBSC_2764; Z1278	Niger	1963	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13	F1-5
52	243	B362; NIBSC_2779; Z1362	Cameroon	1966		Neisseria meningitidis	A	4	ST-4 complex	7	13	F1-5
61	393	B392; NIBSC_2823; Z1392	Greece	1968	carrier	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
64	254	B439; NIBSC_2812; Z1439	Djibouti	1966	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F1-7
67	S5611	B466; NIBSC_2765; Z1466	Australia	1977	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
82	11-004	B503; NIBSC_2826; Z1503	China	1984	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-8
84	IAL2229	B506; NIBSC_2816; Z1506	Brazil	1976		Neisseria meningitidis	A	5	ST-5 complex	20	9	F2-1
90	CN100	B534; NIBSC_2768; Z1534	UK [England]	1941	invasive (unspecified/other)	Neisseria meningitidis	A	21		5-2	10	F3-9

Analysis tools

🕒 Breakdown:	Fields	Two Field	Combinations	Polymorphic sites	Publication	s Sequence bin	
🛃 Analysis:	BURST	Codons	Gene Presence	Genome Comparato	r BLAST	rMLST species id	PCR
Export:	Dataset	Contigs	Sequences				
🗹 Third party:	GrapeTr	ee iTOL	PhyloViz Mic	roreact			

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 and various assembly stats of the sequence bin of each isolate. Note that selecting these will make the analysis run much slower since each contig needs to be examined.

Click submit.

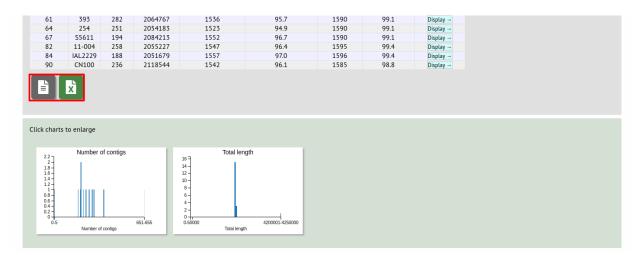
BIGSdb			
Home > Organisms > Neisseria spp.	> Neisseria isolates > Plugins > Sequence bin	breakdown	
Breakdown of sec	luence bin contig prop	erties	Help 🗹 🗿
Please select the required isolate id	s for comparison - use Ctrl or Shift to make multij	ple selections. Select loci/schemes to use	for calculating percentage of alleles designated or tagged.
	Loci	Recommended schemes	Schemes
1 2 7 10 11 13 10 Clear List all	All None Paste list	Select one or more schemes below or use the full schemes list. MLST Ribosomal MLST N. meningitdis cgMLST v1.0 N. gonorrhoeae cgMLST v1.0 CLear	Bexsero Antigen Sequence Human-restricted Neisseri N gongrtheeae coMIST v1 N GMAST NOM V peptide typing Ribosomal MLST eMLST (20 locus partial ge eMLST (20 locus whole gel v
Options	Filter by		Action
Contig analysis (min, max, N5) Calculate %GC	0 etc.) Sequence method: Project: Experiment: 🗸 🕥	~ (j)	RESET SUBMIT

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.

	Clear List all		'16S_rDNA 16S_rRNA (: 23S_rRNA abcZ abcZ (NEIS1 aceF (NEIS1 aceF (NEIS1	.015) .279)		Select one or more sci below or use the full s MLST Ribosomal MLST N. meningitidis cgML N. gonorrhoeae cgM	chemes list. ST v1.0 _ST v1.0	G G G Li M M M P	apsule enetic Information Processin- enomic Islands neage Schemes etabolism gonorrhoeae AMR	
Optio	ns		Fi	lter by			/	Action		
	ig analysis (min ulate %GC	, max, N5	0 etc.)	Sequence method: Project: Experiment:	` (j)	* (i)	* (j	RESET S	UBMIT	
Loci selecto	ed: 1605									
Isolate id	+ Isolate + G	ontigs ÷	Total length +	Alleles designated +	% Alleles desig	nated + Loci tagged	🗧 % Loci tagge	ed 🗧 Sequence bi	n 🕈	
1	A4/M1027	364	2069108	1516	94.5	1591	99.1	Display →		
2	120M	359	2059411	1523	94.9	1593	99.3	Display -		
7	7891	199	2057385	1565	97.5	1594	99.3	Display →		
10	6748	652	4241338	1566	97.6	1590	99.1	Display →		

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.



Sequence bin records can also be accessed by clicking the 'Display' button for each row of the table.

				Experiment:	~ (i)		U	
selecte	d: 1605							
late id 🕈	Isolate +	Contigs +	Total length +	Alleles designated +	% Alleles designated +	Loci tagged 🕈	% Loci tagged 🕈	Sequence bin +
1	A4/M1027	364	2069108	1516	94.5	1591	99.1	Display →
2	120M	359	2059411	1523	94.9	1593	99.3	Display -
7	7891	199	2057385	1565	97.5	1594	99.3	Display -
10	6748	652	4241338	1566	97.6	1590	99.1	Display →
11	129E	272	2072690	1539	95.9	1590	99.1	Display -
13	139M	293	2141711	1555	96.9	1593	99.3	Display -
19	S3131	173	2061338	1561	97.3	1595	99.4	Display -
24	S4355	198	2071992	1557	97.0	1592	99.2	Display →
30	14	1	2145295	1547	96.4	1547	96.4	Display -
31	10	275	2059146	1546	96.3	1592	99.2	Display -
34	20	213	2067373	1552	96.7	1589	99.0	Display -
35	26	194	2056454	1563	97.4	1594	99.3	Display →
52	243	229	2057666	1560	97.2	1594	99.3	Display →
61	393	282	2064767	1536	95.7	1590	99.1	Display →
64	254	251	2054183	1523	94.9	1590	99.1	Display -
67	S5611	194	2084213	1552	96.7	1590	99.1	Display -
82	11-004	258	2055227	1547	96.4	1595	99.4	Display →
84	IAL2229	188	2051679	1557	97.0	1596	99.4	Display -
90	CN100	236	2118544	1542	96.1	1585	98.8	Display -

14.16 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 accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb Home > Organisms > Neisseria spp. > Neisseria isolates *Neisseria* isolates database DI LOG IN Query database **1** SUBMISSIONS Search database Browse, search by any criteria, or enter list of attributes. **PROJECTS EXPORT** + Search by combinations of loci MALYSIS This can include partial matches to find related isolates. CUSTOMISE + **()** INFORMATION **S** TYPING

Jump to the 'Breakdown' category, follow the link to Two Field Breakdown, then click 'Launch Two Field Breakdown'.

The analysis can be selected for the whole database by clicking the 'Two field breakdown' link on the main contents page.

Two Field Breakdown
Summary: Breakdown of one field against another
The two field breakdown plugin generates a table breaking down the frequencies of one field against another. This is analagous to a spreadsheet pivot table. Any primary metadata field, locus, or scheme field can be used and the output can be exported as an Excel spreadsheet.
Documentation bigsdb.readthedocsio 🗗
BCCSS Two 3-Security Nature age - Nature ag
Chars Chars
Launch 'Two Field Breakdown'

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.

										-		
	0090/89				invasive (unspecified/other)					7	16	
13	139M	B99; NIBSC_2795; Z1099		1968		Neisseria meningiti			ST-1 complex	5-2	10	F5-1
	0120/95				invasive (unspecified/other)	Neisseria meningiti					14	
15	1			1999	carrier	Neisseria meningiti						
16	2			1999	carrier	Neisseria meningiti			ST-18 complex			
17	3			1999	carrier	Neisseria meningiti			ST-174 complex			
8	4			1999	carrier	Neisseria meningiti			ST-18 complex	_		
9	S3131	B213; NIBSC_2813; Z1213			invasive (unspecified/other)	Neisseria meningiti			ST-4 complex	7	13-1	F1-5
20	5		,	1999	carrier	Neisseria meningiti			ST-198 complex			
21	6			1999	carrier	Neisseria meningiti			ST-198 complex			
22	7		,	1999	carrier	Neisseria meningiti			ST-60 complex			
23	8			1999	carrier	Neisseria meningiti			ST-32 complex			
	S4355	B227; NIBSC 2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningiti	dis A		ST-5 complex	5-1	9	F3-1
24			-									
	9		Germany	1999	carrier	Neisseria meningiti	dis B	930	ST-334 complex			
es (« Ar	9		Germany	1999	carrier	Neisseria meningiti	dis B	930	ST-334 complex			
s (« Ar	9) (Germany		carrier		dis B	930	ST-334 complex			
s (« Ar	9	own: Fields Two Field)@	Poly		s Sequence bin	PCR	930	ST-334 complex			
s (« Ar	9 alysis 1 Breakdo	own: Fields Two Field	Combinations	Poly	morphic sites Publication	s Sequence bin		930	ST-334 complex			

Select the two fields you wish to breakdown and how you would like the values displayed (percentage/absolute values and totaling options).

BIGSdb				
Home > Organisms > Neisseria spp.	> Neisseria isolates > Plugins > Two field breakdown			
Two field breakdo	own of dataset			Help 🔽
Here you can create a table breakin Isolates	g down one field by another, e.g. breakdown of serogroup by year. Select fields Field 1: clonal complex (MLST) Field 2: capsule group Action RESET SUBMIT	Display • values only values and percentages percentages only	Calculate percentages by dataset row column	

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.

BIGSdb

Show Values a										
Values a		_								
values a	nd pe	rcent	tages							
		-				ule_g			1.	•
olex (MLST) 🔹	A ≎ 1			_	NG \$	W \$	X ¢ Y 1		cnl 🕈	Total 🕈
vatue complex	1 13	С	1	1	4		1	1		12
complex	15		17							17
complex		5								5
complex					1			2		3
complex				1	1	1				3
complex					1					1
complex		3			2					3
complex					2	2				2
complex complex		1				2				1
complex								1		1
complex					1					1
complex		2			1					3
complex		6								6
complex		1								1
complex					1					1
omplex	8				1					8
4 complex		5			1					6
complex complex	6	T								6
complex	0				1				1	2
				1						
complex otal	28	27		1		3	1	3 1	1	1 99

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.

14.17 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 accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb			
Home > Organis	ms > Neisseria spp. > Neisseria isolates		
Neisser	<i>ia</i> isolates database		
	Query database	⇒ LOG IN	
	Search database		
	Browse, search by any criteria, or enter list of attributes.	E PROJECTS	+
		⊡ ª EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.	MALYSIS	
			+
			+

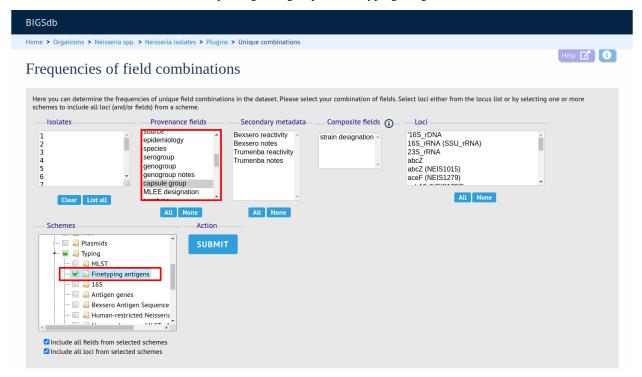
Jump to the 'Breakdown' category, follow the link to Unique Combinations, then click 'Launch Unique Combinations'.

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.

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.

10	TDAM	RAA! INIR207142; 5104A	Philippines	1309		iveisseria meningit	ICIIS	А	T	SI-1 complex	D-7	10	FD-1
4	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningit	idis	Х	117			14	
.5	1		Germany	1999	carrier	Neisseria meningit	idis	E	864				
16	2		Germany	1999	carrier	Neisseria meningit		В	854	ST-18 complex			
17	3		Germany	1999	carrier	Neisseria meningit		W	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningit		В	19	ST-18 complex			
19	S3131	B213; NIBSC_2813; Z1213			invasive (unspecified/other)	Neisseria meningit		А	4	ST-4 complex	7	13-1	F1-5
20	5		Germany	1999	carrier	Neisseria meningit		NG	198	ST-198 complex			
21	6		Germany	1999	carrier	Neisseria meningit		NG		ST-198 complex			
22	7		Germany	1999	carrier	Neisseria meningit		E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningit	idis	В	32	ST-32 complex			
24	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningit	idis	Α	5	ST-5 complex	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria meningit	idis	В	930	ST-334 complex			
, («	00	1 2 3 4 5 6	ക										
(« Ar	alysis 1	tools)		Common bin							
(« Ar	alysis f	tools own: Fields Two Field	Combinations		morphic sites Publication	s Sequence bin							
(« Ar	alysis 1	tools own: Fields Two Field	Combinations Gene Presence		morphic sites Publication me Comparator BLAST	s Sequence bin rMLST species id	PCR						
(« Ar	alysis f	tools own: Fields Two Field yssis: BURST Codons					PCR						

Select the combination of fields to analyse, e.g. serogroup and finetyping antigens.



Click submit. The job will be submitted to the job queue. Once analysis has completed, you will be able to download the results in tab-delimited text or Excel formats.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Job status viewer
Job status viewer
Status Job id: BIGSdb_003448_1596272602_28728 Submit time: 2020-08-01 09.03:22 Status: finished Status: finished Status: 202-08-01 09.03:26 Progres: 10% Stop time: 2020-08-01 09.03:28 Total time: 2 seconds
Output Number of unique combinations: 105 Files
Combinations table (text)
Please note that job results will remain on the server for 7 days.

CHAPTER 15

Data export plugins

15.1 Contig export

The contig export plugin can be accessed by expanding the 'Export' section and clicking the 'Contigs' link in the contents page of isolate databases.

BIGSdb			
Home > Organisms > Neisseria spp.	> Neisseria isolates		
<i>Neisseria</i> isolates	database		
	Query database	DILOG IN	
Search database			
Browse, search by a	any criteria, or enter list of attributes.	PROJECTS	+
		EXPORT	-
Search by combi This can include p	inations of loci artial matches to find related isolates.	Export dataset Contigs Sequences	
		MALYSIS	
			+
		INFORMATION	+

Alternatively, it can be accessed following a query by clicking the 'Contigs' button in the Export section at the bottom of the results table.

					,			-				
11	129E	B92; NIBSC_2828; Z1092	Germany		invasive (unspecified/other)	5	A	1		5-2	10	F3-6
13	139M	B99; NIBSC_2795; Z1099	Philippines	1968		Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
24	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	5-1	9	F3-1
30	14	BennettTree10; alpha14	Germany	1999	carrier	Neisseria meningitidis	cnl	53	ST-53 complex	7	30-3	F5-5
31	10	B269; NIBSC_2825; Z1269	Burkina Faso	1963	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
34	20	B275; NIBSC_2767; Z1275	Niger	1963	invasive (unspecified/other)	Neisseria meningitidis	A			5-2	10	F1-7
35	26	B278; NIBSC_2764; Z1278	Niger	1963	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13	F1-5
52	243	B362; NIBSC_2779; Z1362	Cameroon	1966		Neisseria meningitidis	A	4	ST-4 complex	7	13	F1-5
61	393	B392; NIBSC_2823; Z1392	Greece	1968	carrier	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
64	254	B439; NIBSC_2812; Z1439	Djibouti	1966	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F1-7
67	S5611	B466; NIBSC_2765; Z1466	Australia	1977	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
82	11-004	B503; NIBSC_2826; Z1503	China	1984	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-8
84	IAL2229	B506; NIBSC_2816; Z1506	Brazil	1976		Neisseria meningitidis	A	5	ST-5 complex	20	9	F2-1
90	CN100	B534; NIBSC 2768; Z1534	UK [England]	1941	invasive (unspecified/other)	Neisseria meningitidis	A	21		5-2	10	F3-9

Analysis tools	
----------------	--

🕒 Breakdown:	Fields	Two Field	Combinations	Polymorphic sites	Publication	s Sequence bin	
🛃 Analysis:	BURST	Codons	Gene Presence	Genome Comparato	r BLAST	rMLST species id	PCR
Export:	Dataset	Contigs	Sequences				
🗹 Third party:	GrapeTre	ee iTOL	PhyloViz Mi	croreact			

Select the isolates for which you wish to export contig data for. In databases with a large number of isolates you will need to enter the id numbers rather than select from a list. If the export function was accessed following a query, isolates returned in the query will be pre-selected.

BIGSdb			
Home > Organisms > Neisser	ia spp. > Neisseria isolates > Plugins > Contigs		
Contig analysi	s and export		Help 🔽 🚺
	plate ids from which contigs are associated - use Ctrl or Shift to make bci tagged within the contig - if these loci overlap then the total tagge		
Isolates	Options	Filter by	
1	▲ Identify contigs with >= 0	Sequence method:	
7	FASTA header line: original designation 🗸 👔	Project:	~ (i)
10 11		Experiment: 🗸 👔	
13	*	Minimum length:	
10	Action		
Clear List all	RESET		

At its simplest, press submit.

A table will be produced with download links. Clicking these will produce the contigs in FASTA format.

11 13 19			•				١	finimum length:	` ()		
19			11	Action							
	Clear	List all		RESET	SUE	міт					
Cor	ntigs wi	th >=09	% sequ	ence lens	gth unta	igged					
			matchin	g contigs 🔹	non-matc	hing contigs +					
id ÷	isolate 🕈	contigs ¢	count +	download +	count +	download +					
1	A4/M1027	364	364	*	0						
2	120M	359	359	444444444444444	0						
7	7891	199	199	*	0						
10	6748	652	652	*	0						
11	129E	272	272	*	0						
13	139M	293	293	*	0						
19	S3131	173	173	*	0						
24	S4355	198	198	*	0						
30	14	1	1	*	0						
31	10	275	275		0						
34	20	213	213	-	0						
35	26	194	194	*	0						
52	243	229	229	-	0						
61	393	282	282	-	0						
64	254	251	251		0						
67	S5611	194	194	-	0						
82	11-004	258	258	-	0						
84 90	IAL2229 CN100	188 236	188 236	-	0						
90	CINIDO	200	230	1 20	0						
E	X										

You can also download all the data in a tar file by clicking the 'Batch download' link.

11 13			Ţ				Minimum length:	• (i)	
10			•	Action					
	Clear	List all		RESET	SU	міт			
Co	ntigs wi	th >=09	% sequ	ience leng	gth unta	igged			
14.4	isolate +	contine A	matchi	ng contigs 🔹	non-matc	hing contigs +			
ia ₹	isolale 🕈	contigs =	count +	download 🗢	count +	download +			
1	A4/M1027	364	364	*	0				
2	120M	359	359	*	0				
7	7891	199	199	10 10 10 10 10 10 10 10 10 10 10 10 10 1	0				
10	6748	652	652	*	0				
11	129E	272	272	*	0				
13	139M	293	293	*	0				
19	S3131	173	173	*	0				
24 30	S4355 14	198 1	198 1		0				
31	14	275	275		0				
34	20	213	213	-	0				
35	26	194	194	*	0				
52	243	229	229	*	0				
61	393	282	282	*	0				
64	254	251	251	*	0				
67	S5611	194	194	*	0				
82	11-004	258	258	*	0				
84	IAL2229	188	188	*	0				
90	CN100	236	236	*	0				
_									
	X								
	X	•							
_	_								

15.1.1 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 >= $50 \sim \%$ of sequence untagged
FASTA header line: original designation 🗸 👔

The resulting table has two download links for each isolate, one for contigs matching the condition, and one for contigs that don't match.

11 13 19			•	Action		Minimum length: 🔍 🗸 🕥	
	Clear	List all	[RESET	SUB	T	
Cor	ntigs wi	th >=50	-		ength unta	<u> </u>	
id ÷	isolate 🕈	contigs ¢	⊢		non-matchi		
	4444027	7(4		download		wnload +	
1	A4/M1027 120M	364 359	141 60	1 - 1 - 1 - 1 - 1 - 1	223 299		
7	7891	199	38	1	161		
10	6748	652	377		275	X	
11	129E	272	55	*	217	I I I I I I I I I I I I I I I I I I I	
13	139M	293	164	*	129	I I I I I I I I I I I I I I I I I I I	
19	S3131	173	35	*	138	1 I I I I I I I I I I I I I I I I I I I	
24	S4355	198	45	*	153	±	
30	14	1	0		1	±	
31	10	275	52	*	223	±	
34	20	213	34	*	179	*	
35	26	194	32	*	162	A	
52	243	229	45	.	184	*	
61	393	282	49	*	233	<u>*</u>	
64	254	251	47	* *	204	<u>*</u>	
67	S5611	194	42	-	152	<u> </u>	
82 84	11-004 IAL2229	258	37	-	221	<u>*</u>	
90	CN100	188 236	30 46		158		
90	CINIDO	200	40	- 1	190	·	
C	X						

15.2 Isolate record export

You can export the entire isolate recordset by expanding the Export section on the main contents page and clicking the 'Export dataset' link.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria isolates		
<i>Neisseria</i> isolates database		
Query database	DIG IN	
Search database		
Browse, search by any criteria, or enter list of attributes.	PROJECTS	+
	EXPORT	-
Search by combinations of loci This can include partial matches to find related isolates.	Export dataset Contigs Sequences	
	MALYSIS	
		+
	1 INFORMATION	+

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.

.6	2			Germany		carrie		Neisseria meningi		В		ST-18 complex			
7	3			Germany	1999	carrie	er 1	Neisseria meningi	tidis	W	174	ST-174 complex			
8	4			Germany	1999	carrie	er 1	Neisseria meningi	tidis	В	19	ST-18 complex			
	S3131	B213; NIBSC_2	2813; Z1213	Ghana		invasive (unspe		Neisseria meningi		A	4	ST-4 complex	7	13-1	F1-
)	5			Germany	1999	carrie	er 1	Neisseria meningi	tidis	NG	198	ST-198 complex			
	6			Germany	1999	carrie	er 1	Neisseria meningi	tidis	NG	198	ST-198 complex			
	7			Germany		carrie		Neisseria meningi		E	60	ST-60 complex			
5	8			Germany	1999	carrie	er 1	Neisseria meningi	tidis	В	32	ST-32 complex			
	S4355	B227; NIBSC_2	2806; Z1227	Denmark	1974	invasive (unspe	cified/other) 1	Neisseria meningi	tidis	Α	5	ST-5 complex	5-1	9	F3-
5	9			Germany	1999	carrie	er 1	Neisseria meningi	tidis	В	930	ST-334 complex			
×)()	123	4 () ()											
« An) 🕢 alysis t	123 tools	4	00											
	-	1 2 3 tools own: Fields	4	Combinatio	ns Poly	morphic sites	Publications	Sequence bin							
	-	own: Fields	4 Two Field Codons	Combinatio Gene Presence		morphic sites		Sequence bin rMLST species id	PCR						
	Breakdo	own: Fields	Codons					· · ·	PCR						

Select the isolate ids (if they have not been pre-selected from your query), isolate fields and schemes to include.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Export dataset
Export dataset
This script will export the dataset in tab-delimited text and Excel formats. Select which fields you would like included. Select loci either from the locus list or by selecting one or more schemes to include all loci (and/or fields) from a scheme.
— Isolates — — Provenance fields — — Secondary metadata — Composite fields ① _ References —
1 id id isolate isolate
Schemes Classification schemes Options
Mm_cdgc_200 Indicate sequence status if no allele defined () Mm_cdgc_100 Nm_cdgc_200 Mm_cdgc_200 Include locus common names Mg_cdgc_300 Ng_cgc_300 Ng_cdgc_200 Molecular weights Antigen genes CfG/TTG at start codes for methionine
☑ Include all fields from selected schemes ☑ Include all loci from selected schemes

Click Submit.

You can then download the data in tab-delimited text or Excel formats.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Export dataset	
Export dataset	Help 🔽
Please wait for processing to finish (do not refresh page).	
Output files being generated done	

Export jobs for larger datasets will be sent to the job queue.

15.2.1 Advanced options

Options

Indicate sequence status if no allele defined (i)

Include locus common names

Export allele numbers

Use one row per field

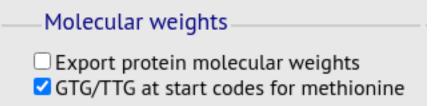
- Include isolate field in row (used only with 'one row' option)
- Export full allele designation record (used only with 'one row' option)

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.

15.2.2 Molecular weight calculation

The plugin can also calculate the predicted molecular weight of the gene product of any allele designated in the dataset.



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.

15.3 Profile export

You can export the allelic profiles for any indexed scheme (those containing a primary key field) defined in the sequence definition database.

The profile export function can be accessed by expanding the 'Export' section and clicking the 'Profiles' link on the contents page.

BIGSdb

Home > Organisms > Organism > Neisseria typing

Neisseria typing database

Query a sequence	Find alleles	Search for allelic profiles	DI LOG IN	
Single sequence	By specific criteria	By specific criteria		
Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	L DOWNLOADS	+
induires.			EXPORT	-
Batch sequences Query multiple independent sequences in FASTA format to identify allelic matches.	By locus Select, analyse and download specific alleles from a single locus.	By allelic profile This can include partial matches to find related profiles.	Locus sequences Profiles Profile sequences	
		In a batch	MALYSIS	+
		Look up multiple allelic profiles together.		+
			B INFORMATION	+
			S ISOLATES	

Alternatively, you can access this function by clicking the 'Profiles' button in the Export list at the bottom of a results table. Please note that the list of functions here may vary depending on the setup of the database.

		±~	**	'	~	**		51 207 comptex	
14	4	1	15	7	8	11	1	ST-269 complex	
15	13	3	16	1	3	11	9	ST-364 complex	
16	15	9	9	13	8	19	15	ST-4240/6688 complex	
17	8	3	13	1	11	12	4		
18	7	8	10	19	10	1	2	ST-18 complex	
19	7	8	10	19	8	1	2	ST-18 complex	
20	6	8	10	17	10	1	2	ST-18 complex	
21	1	5	1	1	2	16	17		
22	11	5	18	8	11	24	21	ST-22 complex	
23	10	5	18	9	11	9	17	ST-23 complex	
24	2	5	2	7	15	20	5	ST-750 complex	
25	6	5	2	12	6	13	14		
(3) (25 6 5 2 12 6 13 14) (3) (3) (3) (3) (3) (3) (3) (3) (3) (3									
Analysis tools									

Export: Profiles Sequences

This will take you to a form with a list box in which the identifiers of the profile definitions you wish to include can be entered. Following a query, these values will be pre-entered. If the box is left empty then all profiles will be included. You can optionally include provenance information (sender, curator and datestamps) by selecting the appropriate checkboxes.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Export allelic profiles - N		Help 🗹
Schemes Please select the scheme you would like to query:		
MLST v Se		
This script will export allelic profiles in tab-delimiteSelect STs	ed text and Excel formats. Provenance	
Paste in list of ids to include, start a new line for each. Leave blank to include all ids.	Include sender details Include curator details Action SUBMIT	

Click submit.

The export job will be submitted to the job queue.

BIGSdb
Home > Organisms > Organism > Neisseria typing > Job status viewer
Job status viewer
Status Job id: BIGSdb_022243_1596292027_18737 Submit time: 2020-08-01 14:27.07 Status: finished Start time: 2020-08-01 14:27:21 Progres: 100% Stap time: 2020-08-01 14:27:33 Total time: 12 seconds
Output Files Profiles - Tab-delimited text (text) Profiles - Excel format (Excel)
Please note that job results will remain on the server for 7 days.

The profiles will be exported in tab-delimited text and Excel formats.

15.4 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 expanding the 'Export' section and clicking the 'Sequences' link on the contents page of isolate databases, or the 'Profile sequences' link on the contents page of sequence definition databases.

BIGSdb			
	ms > Neisseria spp. > Neisseria isolates		
Neissei	<i>ia</i> isolates database		
	Query database	DI LOG IN	
	Search database		
	Browse, search by any criteria, or enter list of attributes.	PROJECTS	+
		EXPORT	-
	Search by combinations of loci This can include partial matches to find related isolates.	Export dataset Contigs Sequences	
		MALYSIS	
		¢ CUSTOMISE	+
			+

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.

	~			Germany	±///		~ .	THE SACTOR THE TH	MINIMUM			21 27 1 2011/2020			
18	4			Germany	1999	carri	er	Neisseria menir		В	19	ST-18 complex			
19	S3131	B213; NIBSC_2	813; Z1213	6 Ghana	1973	invasive (unspe	cified/other)	Neisseria menir	gitidis	А	4	ST-4 complex	7	13-1	F1-5
20	5			Germany	1999	carri	er	Neisseria menir	gitidis	NG	198	ST-198 complex			
21	6			Germany	1999	carri	er	Neisseria menir	gitidis	NG	198	ST-198 complex			
22	7			Germany	1999	carri	er	Neisseria menir		E	60	ST-60 complex			
23	8			Germany	1999	carri		Neisseria menir		В	32	ST-32 complex			
24	S4355	B227; NIBSC_2	806; Z1227	Denmark		invasive (unspe	cified/other)	Neisseria menir		A	5	ST-5 complex	5-1	9	F3-1
25	9			Germany	1999	carri	er	Neisseria menir	gitidis	В	930	ST-334 complex			
Ċ		123	* 0	90											
Ar	nalysis t		* (00											
	nalysis t Breakdo	tools	Two Field	Combinations	Polyr	morphic sites	Publication	s Sequence bi	1						
		tools	Two Field Codons	Combinations Gene Presence		morphic sites me Comparator		s Sequence bi rMLST species i							
	Breakdo	tools own: Fields bysis: BURST	Codons			- · · · · · · · · · · · · · · · · · · ·									

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.

1 1		
files. Only DNA loci that have a corresponding data that you would like to include. Alternatively select replaced with gap characters. Aligned output is limited to 200 records; total outp	base containing allele sequence identifiers, or DNA and peptid	rty applications, such as ClonalFrame. It will also produce concatenated FASTA e loci with genome sequences tagged, can be included. Please check the loci se scheme. If a sequence does not exist in the remote database, it will be
Select ids	Include in identifier	Loci
Paste in list of ids to include, start a new line for each. Leave blank to include all ids.	isolate country region year date sampled isoyear sampled week sampled date received non culture	'16S_rDNA 16S_rRNA (SSU_rRNA) 235_rRNA abcZ abcZ (NEIS1015) acber (NEIS1279) All None Paste list
Schemes	Options	Action
Plasmids Typing MLST Grinetyping antigens 165 Grinetyping genes	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 Do not include sequences with problem flagged (define Do not include incomplete sequences	SUBMIT
🔲 🚚 Bexsero Antigen Sequence	Include 0 v bp flanking sequence (i)	
🔲 🚚 Human-restricted Neisseria	Align sequences	
	Aligner: MAFFT Translate sequences	

Click submit. The job will be submitted to the job queue.

Sequences will be export in XMFA and FASTA file formats.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Job status viewer
Job status viewer
Status Job Id: BIGSdb_029655_1596295569_17972 Submit time: 2020-08-01 15:26:09 Status: finished Start time: 2020-08-01 15:26:13 Progress: 100% Stop time: 2020-08-01 15:26:19 Total time: 6 seconds
Output Files
XMFA output file (not aligned) Concatenated FASTA (not aligned) Concatenated FASTA (not aligned) Tar file containing all output files
Please note that job results will remain on the server for 7 days.

15.4.1 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
If both allele designations and tagged sequences exist for a locus, choose how you want these handled: (i)
Use sequences tagged from the bin Use allele sequence retrieved from external database
 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 ✓ □ Translate 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.

CHAPTER 16

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.

BIG	Sdb				
Home > Organis	ms > Organism > Neisseria typing				
Neiss	<i>seria</i> typing database				
	Query a sequence	Find alleles	Search for allelic profiles	●〕LOG IN	
	Sequence query	Sequence attribute search	Allelic profile query	1 SUBMISSIONS	
	Query a single sequence or whole genome assembly to identify allelic matches.	Find alleles by matching criteria (all loci together)	Search, browse or enter list of profiles	Ł DOWNLOADS	+
				Z EXPORT	+
	Batch sequence query Query multiple independent sequences	Locus-specific sequence attribute search	Search by combinations of alleles This can include partial matches to find	MALYSIS	+
	in FASTA format to identify allelic matches.	Select, analyse and download specific alleles from a single locus.	related profiles.	CUSTOMISE	+
			Batch profile query	INFORMATION	+
			Lookup multiple allelic profiles together.	SISOLATES	

16.1 Registering a user account

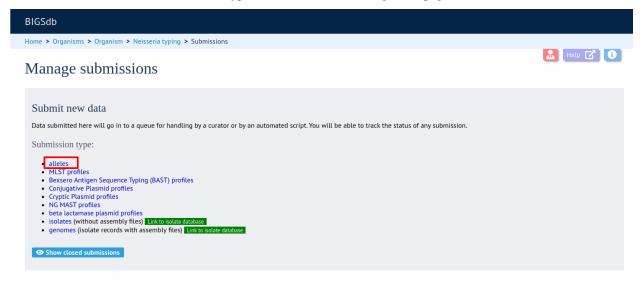
You must have an account for the appropriate database in order to use the submission system. On systems utilizing site-wide databases, such as PubMLST, this can be done automatically via the web. Other sites may require you to contact a curator to set this up.

16.2 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.

16.2.1 Start

Click the 'alleles' link under submission type on the submission management page.



16.2.2 Select the submission locus

Select the locus from the locus list box:

BIGSdb		
Home > Organisms > Organism > Neisseria typing	Submissions > New submission	
Manage submissions		🔝 Heip 🗹 🚺
	locus for which you have new alleles - this is because diff nould be trimmed to the correct start/end sites for the sel	erent loci may have different curators. You can submit any number of new sequences
Filter loci by scheme	Select locus Filter 165_rDNA 235_rRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) accA2 (NEIS1727) acnA (NEIS1729) acnB (NEIS1492) FASTA or single sequence	Sequence details technology:1 Illumina read length:1 200-299 coverage:1 20-498 × assembly: de novo × assembly: de novo × assembly: software:1 Velvet Sequence length outside usual range Action

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'.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Submiss	ions > New submission	
Manage submissions		🔝 Help 🗹 🚯
Submit new alleles You need to make a separate submission for each locus for whi for a single locus as one submission. Sequences should be trim		ay have different curators. You can submit any number of new sequences
Filter loci by scheme	Select locus	Sequence details
All loci Strain typing Strain typing Genetic Information Processing Genetic Islands	Filter 165_rDNA 235_rRNA abcZ abcZ (NEIS1015) accF (NEIS1279) ackA2 (NEIS1727) acnA (NEIS1729) acnB (NEIS1729) acnB (NEIS1492)	technology: I Illumina read length: 200-299 coverage: 20-49x assembly: de novo assembly: de novo assembly: software: Velvet ⊆ Sequence length outside usual range ①
🔲 🚚 N. gonorrhoeae AMR	FASTA or single sequence	Action
 Plasmids Typing antigens Typing antigens Typic species Typic spe		SUBMIT
- Question		
 OMV peptide typing 		
- I III III IIII IIII IIIIIIIIIIIIIIII		

The locus list is now constrained making selection easier.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Submissio	ns > New submission	
Manage submissions		2
Submit new alleles You need to make a separate submission for each locus for whic for a single locus as one submission. Sequences should be trimm	h you have new alleles - this is because different loci may have different curators. You can submit any number of new sequences ned to the correct start/end sites for the selected locus.	
Filter loci by scheme	Select locus Sequence details Filter abcZ adk aroE fumC gdh pdhC pgm technology! Illumina ack aroE fumC gdh pdhC pgm coverage! 200-299 \cdot assembly: de novo \cdot gdh pdhC pgm ssembly: de novo \cdot	
Metabolism Metabolism M. gonorrhoeae AMR Plasmids Mist Mist	FASTA or single sequence Action	

16.2.3 Enter details of sequencing method

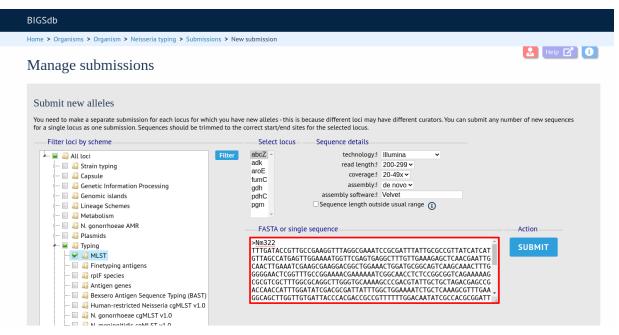
There are a number of fields that must be filled in so that the curator knows how the sequence was obtained:

- technology the sequencing 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.

16.2.4 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.



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.

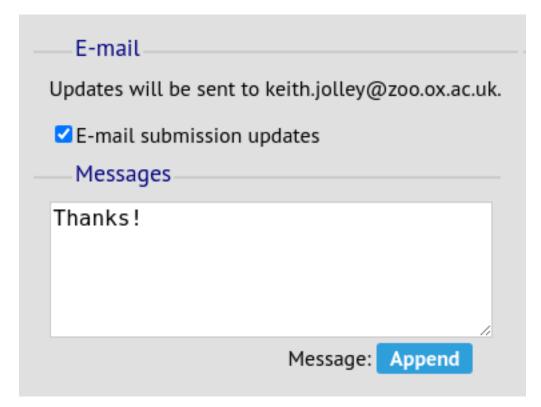
BIGSdb			
Home > Organisms > Organism > Neisseria typing	> Submissions > New submission		
Manage submissions			
Error: Sequence 'Nm324' has already been det Submit new alleles You need to make a separate submission for each la		zcause different loci may have different curators. You can submit an	y number of new sequences
for a single locus as one submission. Sequences sho — Filter loci by scheme — 🔊 🚇 All loci	Select locus Filter '16S_rDNA	for the selected locus. Sequence details technology: Illumina	
Strain typing Genetic Information Processing Genetic Information Genetic Information	23S_rRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) ackA2 (NEIS1727)	read length: 100-199 v coverage: 20-49x v assembly: de novo v assembly software: Spades	
	acnA (NEIS1729) acnB (NEIS1492) FASTA or single sequence	Sequence length outside usual range Action	n
	>Nm322 TTTGATACCGTTGCCGAAGGTTTA GTTAGCCATGAGTTGGAAAATGGT CAACTTGAAATCGAAGCGAAG	AGGCGAAATCCGCGATTTATTGCGCCGTTATCATCAT TTCGAGTGAGGCTTTGTTGAAAGAAGCTCAACGAATTG CGGCTGGAAACTGGATGCGGCAGTCAACCAAACTTTG CGACAAAATCGGCAACCTCTCCGGCGGTCAGAAAAAG GGTGCAAAAGCCCGACGTATTGCTGCTGAGCAGAGCCG	МІТ
	ACCAACCATTTGGATATCGACGCG	SATTATTTGGCTGGAAAATCTGCTCAAAGCGTTTGAA CGACCGCCGTTTTTGGACAATATCGCCACGCGGATT	

Assuming the preliminary checks have passed you will then be able to add additional information to your submission.

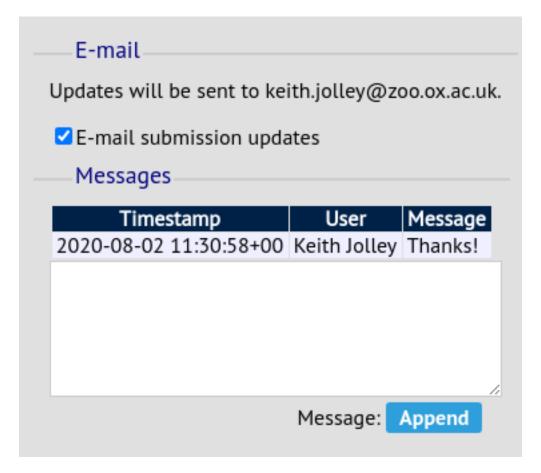
BIGSdb	
Home > Organisms > Organism > Neisseria typing > Submissions > New submission	
Manage submissions	Help 🗹 🕕
Supporting files Please note that if you are submitting alleles determined by Sanger sequencing that you must upload both forward a pack or compress your files (zip, rar etc).	nd reverse trace files. Your submission will be rejected if you don't. Please do not
Abort submission Submission: BIGSdb_20200802122845_002379_40069 Supporting files Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an ex Individual filesize is limited to 64 MB. You can upload up to 64 MB in one go (the upload may fail if you try to do m can upload multiple times to that the total size of the submission can be larger.	
Drop files here or click to	upload.
Sequences	Sequence details
You are submitting the following abcZ sequences: Download This Identifier Length Sequence Status Query Assigned alleld Nm322 455 TTTGATACCGTTGCCGAAGG CCGGATTGTCGAACTCGATC Pending Q Nm323 433 TTTGATACTGTTGCCGAAGG CCGGATTGTCGAACTCGATC Pending Q	technology:1 Illumina read length:1 200-299 ↓ coverage:1 20-49x ↓ assembly:2 de novo ↓ assembly software:1 Velvet Sequence length outside usual range ĵ
E-mail Action	
Updates will be sent to keith.jolley@zoo.ox.ac.uk.	
Message: Append	

16.2.5 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Append'. This is not normally necessary for routine submissions.



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.



16.2.6 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 dragging and dropping in to the large dotted area in the 'Supporting files' section. Alternatively, you can click this area and select files from the local file system.

	oad up to 64 MB in one go (the upload may fail if you try to	an explanatory note so that they can be linked to the appropriate submissio do more than this - just try again with fewer files at a time if it does), altho to upload.	
Nm322 433 TTTGATACCGTTGCCGAAGG	Download FAS Uence Status Query Assigned a GCGGATTGTCGAACTCGATC pending Q GCGGATTGTCGAACTCGATC pending Q	Sequence details technology: Illumina → read length: 200-299 → allele coverage: 20-49x → de novo →	<i>\</i>

The files will be uploaded and shown in a table.

	ration. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. oad up to 64 MB in one go (the upload may fail if you try to do more than this - just try again with fewer files at a time if it does), although you 'the submission can be larger.
	Drop files here or click to upload.
Uploaded files Filename Size Detete NM322_abcZ_Rab1 252 KB - NM322_abcZ_Fab1 269.6 KB - NM322_abcZ_Fab1 281.8 KB - NM322_abcZ_Rab1 258.7 KB - Detete - -	

Files can be removed from the submission by checking the appropriate 'Delete' box and clicking 'Delete selected files'.

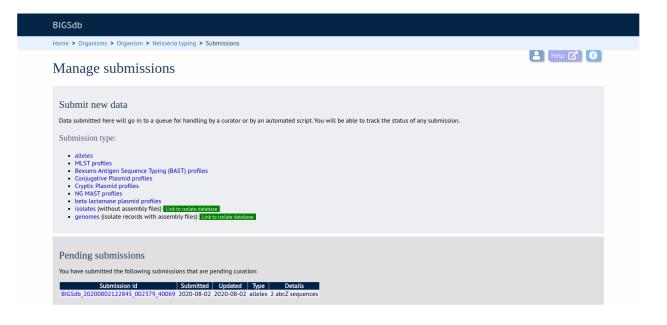
16.2.7 Finalize submission

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!'.

Sequences	Sequence details
You are submitting the following abcZ sequences: Download 🏊	technology: Illumina 🗸
	read length: 200-299 ~
Identifier Length Sequence Status Query Assigned allele	coverage:! 20-49x ✓
Nm322 433 TTTGATACCGTTGCCGAAGG GCGGATTGTCGAACTCGATC pending Q	assembly:! de novo 🗸
Nm323 433 TTTGATACTGTTGCCGAAGG GCGGATTGTCGAACTCGATC pending Q	assembly software: Velvet
	Sequence length outside usual range (i)
E-mail Action	
Updates will be sent to keith.jolley@zoo.ox.ac.uk.	
FINALIZE SUBMISSION!	
E-mail submission updates	
Messages	
Timestamp User Message	
2020-08-02 11:30:58+00 Keith Jolley Thanks!	
Message: Append	
Message. Append	

Your submission will then be listed under 'Pending submissions' on your submission page.

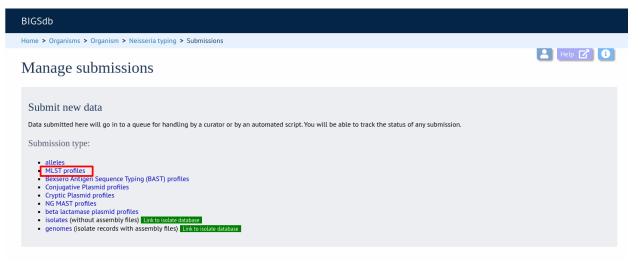


16.3 Profile submission

16.3.1 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.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Submissions > New submission		
Manage submissions		
Submit new MLST profiles Paste in your profiles for assignment using the template available below. Templates		
Please paste in tab-delimited text (Include a field header line)	Action	

16.3.2 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.

BIGSdb				
Home > Organisms > Organism > Neisseri	a typing > Submissions	> New submission		
Manage submissions				Help 🕜 🚯
Submit new MLST profiles Paste in your profiles for assignment using Templates Please paste in tab-delimited text id <u>abcZ</u> <u>adk</u> <u>aroE</u> 5 7 3 3 9 6 9			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.

BIGSdb										
Home > Or	ganisms > C)rganism:	> Neisseria	typing > S	ubmissions	> New sub	mission			
Mana	ige sul	omis	sions							🛓 Help 📝 🚯
×	Error: Row 4: Pro	file has alr	eady been c	lefined as S	T-44.					
	t new M	-		he template	available	below.				
Templ	X					- 11 2				
Plea	se paste in abcZ	adk	aroE	fumC	gdh	r une) pdhC	pgm		Action	
10	8 5 3 9	5 7 6 6	32 3 4 9	2 6 3 9	6 33 8 9	43 12 4 6	2 32 6 9		SUBMIT	

Provided the checks pass, you will then be able to add additional information to your submission. New profile submissions usually don't require supporting files directly in the submission. You generally will need to make a corresponding *submission to the isolate database* though.

16.3.3 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Append'.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Submissions > New submission	n	
Manage submissions		
Isolate submission required Please note that you must make a corresponding submission <i>to the isolate database</i> details. Every newly defined ST should have a representative isolate record in the iso		
Abort submission Submission: BIGSdb_20200802163741_003907_56760 Supporting files Please upload any supporting files required for curation. Ensure that these are nan Individual filesize is limited to 64 MB. You can upload up to 64 MB in one go (the u can upload multiple times so that the total size of the submission can be larger.	ned unambiguously or add an explanatory note so tha	
Drop files	s here or click to upload.	
Profiles	E-mail	Action
You are submitting the following MLST profiles: Download cov Identifier abc2 adk aroE [rumC] gdh pdhC pgm Status Query Assigned ST Row 1 8 5 32 2 6 43 32 pending Q Row 2 5 7 3 6 33 12 2 pending Q Row 3 5 6 4 3 8 4 6 pending Q	Updates will be sent to keith.jolley@zoo.ox.ac.uk.	FINALIZE SUBMISSION!
Messages Corresponding isolate submission has been made. <u>BIGSdb</u> 20200802163741_003123_46421 Message: Append		

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.

Messages		
Timestamp	User	Message
2020-08-02 15:40:36+00	Keith Jolley	Corresponding isolate submission has been made. BIGSdb_20200802163741_003123_46421
		Message: Append

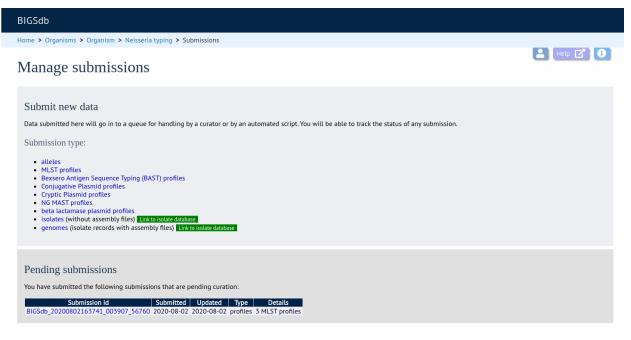
16.3.4 Finalize submission

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!'.



Your submission will then be listed under 'Pending submissions' on your submission page.



16.4 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.

16.4.1 Start

Click the 'isolates' link under submission type on the submission management page.



Download the Excel submission template.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Submissions	
Manage submissions	Help 🗹 🕚
Submit new isolates Paste in your isolates for addition to the database using the template available below. Optionally enter aliases (alternative names) for your isolates as a semi-colon (;)-separated list. Optionally enter references for your isolates as a semi-colon (;)-separated list of PubMed ids. Out can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name Description Please paste in tab-delimited text (Include a field header line) Actio	
SUB	міт

16.4.2 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.

anage submissions ubmit new isolates te in your isolates for addition to the database using the template available below. • Optionally enter references for your isolates as a semi-colon ()-separated list. • Optionally enter references for your isolates as a semi-colon ()-separated list. • You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name. • Optionally enter references Country region year date_sampled salate received non_culture epidemiological_year age_yr age_range Action salate received non_culture epidemiological_year age_yr age_range Subphonamide referinsers Country region year date_sampled salate received non_culture epidemiological_year age_yr age_range Subphonamide referinsers Subphonamide ceftriaxone referinsers subphonamide ceftriaxone referinsers anxi:illin subphonamide ceftriaxone referinsers abcz adk aroo func gend referinsers More Mark aroo func gend referinscone subphonamide referinsers abcz adk aroo func gend referinscone reffrinscone rifamp	
te in your isolates for addition to the database using the template available below. • Optionally enter references for your isolates as a semi-colon ()-separated list. • Optionally enter references for your isolates as a semi-colon ()-separated list of PubMed ids. • You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name. • Optionally enter references for your isolates as a semi-colon ()-separated list of PubMed ids. • You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name. • Optionally enter references country region year date_sampled alate - second year age_yr age_range endemiciol cological year age_yr age_range genogroup genogroup ltEE designation serotype sero subtype ET_no penicillin sulphonamide ceftriaxone range chloramphenicol chloramphenicol range ceftoraxime infamplicin range ciprofloxacin range tetracycline cefixime azithromycin bioproject accession biosample accession bi	
 Optionally enter aliases (alternative names) for your isolates as a semi-colon ()-separated list. Optionally enter references for your isolates as a semi-colon ()-separated list of PubMed ids. You can also upload additional altele fields along with the other isolate data - simply create a new column with the locus name. The set of the second secon	
 Optionally enter references for your isolates as a semi-colon ()-separated list of PubMed lds. You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name. Image: Column Colum	
solate aliases references country region year date_sampled late_received non_culture epidemiological_year age_yr age_range age_mth sex disease source epidemiology species <u>serogroup</u> genogroup lettlef designation <u>serotype</u> sero subtype ET no penicillin ceficiance range chloramphenicol chloramphenicol range ceftriaxone ietefriaxone range chloramphenicol chloramphenicol range ciprofloxacin rifampicin range ciprofloxacin pectinomycin bioproject accession biosample_accession EMA run accession rivate_project comments abcz adk aroE fumC qdh pdhC rivate_project comments abcz PetA VR gyrA penA rpoB NG- NGST NG NG-NAST tbpB K325 UK 2020 meningitis and septicaemia blood Neisseria meningitidis C 2 3 4 3 8 4 6 5 2 F1-5	
solate aliases references country region year date_sampled late_received non_culture epidemiological_year age_yr age_range age_mth sex disease source epidemiology species <u>serogroup</u> genogroup lettlef designation <u>serotype</u> sero subtype ET no penicillin ceficiance range chloramphenicol chloramphenicol range ceftriaxone ietefriaxone range chloramphenicol chloramphenicol range ciprofloxacin rifampicin range ciprofloxacin pectinomycin bioproject accession biosample_accession EMA run accession rivate_project comments abcz adk aroE fumC qdh pdhC rivate_project comments abcz PetA VR gyrA penA rpoB NG- NGST NG NG-NAST tbpB K325 UK 2020 meningitis and septicaemia blood Neisseria meningitidis C 2 3 4 3 8 4 6 5 2 F1-5	
The project comments abcz adk aroE fumC gdh penA rpoB NG- NGST prof NG	
LEE designation serotype sero subtype ET no penicillin renicillin range amoxicillin sulphonamide ceftriaxone renicillin range amoxicillin sulphonamide ceftriaxone reftriaxone range chloramphenicol range cefotaxime reftriaxone rifampicin rifampicin range ciprofloxacin riprofloxacin range tetracycline cefixime azithromycin pectinomycin bioproject accession EMA run_accession orivate_project comments abcz Adk aroE AST porA VR1 PorA VR2 FetA VR gyrA penA rpoB K325 UK 2020 tesseria meningitidis C reningitis and septicaemia blood Neisseria meningitidis C 2 3 4 3 8 4 6 5 F1-5	
reftriaxone range chloramphenicol chloramphenicol range ceftoraxime reftoraxime range rifampicin range ciprofloxacin rifampicin range ciprofloxacin rigentinomycin bioproject accession biosample accession EMA run_accession rivate_project coments abcZ adk arof func gdh pdhC rgm PorA VR1 PorA VR2 FetA VR gyrA penA rpoB NG- K325 WG-MAST_tbpB K325 WG-MAST_tbpB K325 WG blood Neisseria meningitidis C 2 3 4 3 8 4 6 5 2 F1-5	
refotaxime_rangeifampicinifampicin_rangeciprofloxaciniprofloxaciniprofloxacinarangetetracyclinecefiximeazithromyciniprofloxacininspectionswippications	
pectinomycin bioproject accession biosample accession ENA run accession rivate project comments abcz adk aroE fumC gdh pdhC ggm PorA VR1 PorA VR2 FetA VR gyrA penA rpoB NG- HAST_porB NG-MAST_tbpB HX325 UK 2020 heningitis and septicaemia blood Neisseria meningitidis C z 3 4 3 8 4 6 5 2 F1-5	
private_project comments <u>abcZ</u> adk <u>aroE fumC gdh pdhC</u> ggm <u>PorA VR1 PorA VR2 FetA VR gyrA penA rpoB</u> NG- IAST porB <u>NG-MAST tbpB</u> IK325 UK 2020 Meningitis and septicaemia blood <u>Neisseria meningitidis</u> C 2 3 4 3 8 4 6 5 2 F <u>1</u> -5	
NG-MAST_tbpB NG-MAST_tbpB IK325 UK 2020 ieningitis and septicaemia blood Neisseria meningitidis C 2 3 4 3 8 4 6 5 2 F1-5	
IK325 UK 2020 meningitis and septicaemia blood Neisseria meningitidis C 2 3 4 3 8 4 6 5 2 F1-5	
2 3 4 3 8 4 6 5 2 F1-5	
epticaemia blood <u>Neisseria meningitidis</u> W	
2 3 4 3 18 4 6 5-1 2 <u>F1</u> -5	

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.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Submissions	
Manage submissions	Leto 🔽 🚯
Error: UK325 has problems - species: value is not on the list of allowed values for this field.	
Submit new isolates Paste in your isolates for addition to the database using the template available below. • Optionally enter aliases (alternative names) for your isolates as a semi-colon (;)-separated list. • Optionally enter references for your isolates as a semi-colon (;)-separated list of PubMed ids. • You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name.	
Please paste in tab-delimited text (include a field header line) Action isolate aliases references date_received non_culture country region year epidemiological_year date_sampled age_yr age_range SUBMIT	

Provided the checks pass, you will then be able to add additional information to your submission.

16.4.3 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Append'.

BIGSdb	
lome > Organisms > Neisseria spp. > Neisseria isol	
Manage submissions	💄 Help 📝 🔮
New MLST profiles? If your isolate record contains a new combination of your STs will be assigned and that you get notified o	4LST alleles, please also make a new profile submission <i>to the sequence database</i> . Please see the submission guide for details. This ensures that these assignments.
Abort submission Submission: BIGSdb_2020080216 Isolates You are submitting the following isolates: Downlo	
isolate country year disease	source species serogroup abc2 adk aroE fumC gdh pdhC pgm PorA_VR1 PorA_VR2 FetA_VR ST (MLST) blood Neisseria meningitidis C 2 3 4 3 8 4 6 5 2 F1-5 11
UK325 UK 2020 meningitis and septicaemi UK326 UK 2020 septicaemia	blood Neisseria meningitidis C 2 3 4 3 8 4 6 5 2 F1-5 11 blood Neisseria meningitidis W 2 3 4 3 18 4 6 5-1 2 F1-5 -
Messages	E-mail Action
I think <u>UK326</u> has a new <u>MLST</u> profile. I will also submit a new profile submission to the sequence definition database for this. Message: <u>Append</u>	Updates will be sent to keith.jolley@zoo.ox.ac.uk. FINALIZE SUBMISSION! ■ E-mail submission updates

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.

Messages		
Timestamp	User	Message
2020-08-02 15:54:39+00	Keith Jolley	I think UK326 has a new MLST profile. I will also submit a new profile submission to the sequence definition database for this.
		^
		Message: Append

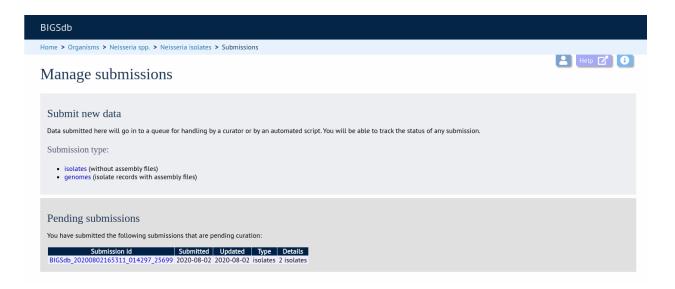
16.4.4 Finalize submission

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!'.



Your submission will then be listed under 'Pending submissions' on your submission page.



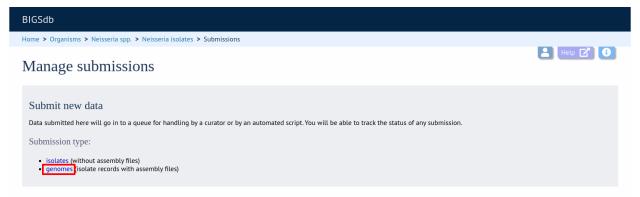
16.5 Genome submission

Submitting genomes uses the same process as standard *isolate submission*. The only difference is that there are a couple of extra required fields in the submission table:

- assembly_filename this is the name of the FASTA file containing the assembly contigs. This must be uploaded
 as a supporting file you will not be able to finalize the submission until every isolate record has a matching
 contig file.
- sequence_method the sequencing technology used to generate the sequences. The allowed values are listed on the submission page.

Locus fields are not usually included in a genome submission as these can be readily extracted from the genome.

To start the submission, click the 'genomes' link under submission type on the submission management page.



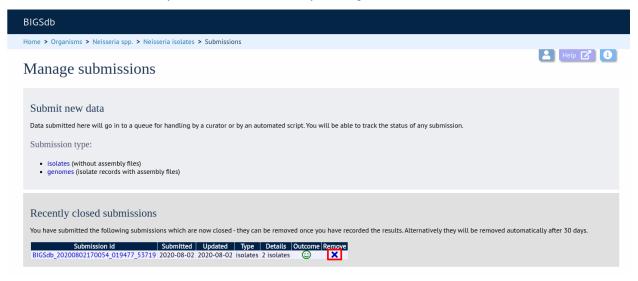
Then follow the steps for *isolate submission*, uploading the contig files as supporting files.

Note: When including the filename for your FASTA file containing the genome assembly, please note that Windows will, by default, hide the file extension, e.g. .fas or .fasta. Even if it is hidden in the Windows interface, the file extension is part of the filename and must be included so that the uploaded file has exactly the same name as entered in the submission template. See https://www.techadvisor.co.uk/how-to/windows/windows-10-file-extensions-3697651

to see how to display hidden file extensions in Windows 10.

16.6 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.



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.

CHAPTER 17

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.

17.1 Passing additional/optional parameters

If you are using a method called with GET, 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.

https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates?page=2&page_size=100

Methods called with POST require their arguments to be sent as JSON within the post body.

17.2 Paging using request headers

Paging of results can be selected using query parameters as described above. In this case, methods that support paging will include a paging object in the JSON response. This will contain links to the next page, last page etc.

The API also supports paging using request headers. The following request headers are supported:

- X-OFFSET
- X-PER-PAGE

e.g.

If either of these headers are used, the paging object is no longer returned as part of the JSON response. The response will include the following headers:

- X-OFFSET
- X-PER-PAGE
- X-TOTAL-PAGES

17.3 Resources

- GET / or /db List site resources
- *GET /db/{database}* List database resources
- *GET /db/{database}/classification_schemes* List classification schemes
- *GET /db/{database}/classification_schemes/{classification_scheme_id}* Retrieve classification scheme information and groups
- *GET /db/{database}/classification_schemes/{classification_scheme_id}/groups/{group_id}* List isolates belonging to a classification scheme group
- 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
- POST /db/{database}/loci/{locus}/sequence Query sequence to identify allele
- POST /db/{database}/sequence Query sequence to identify allele without specifying locus
- GET /db/{database}/sequences Get summary of defined sequences
- GET /db/{database}/schemes List schemes
- GET /db/{database}/schemes/{scheme_id} Retrieve scheme information
- GET /db/{database}/schemes/{scheme_id}/loci Retrieve scheme loci
- 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
- *POST /db/{database}/scheme_id}/sequence* Query sequence to extract allele designations/fields for a scheme
- POST /db/{database}/schemes/{scheme_id}/designations Query allelic profile to extract fields for a scheme
- GET /db/{database}/isolates Retrieve list of isolate records
- GET /db/{database}/genomes Retrieve list of isolate records that have genome assemblies
- *POST /db/{database}/isolates/search* Search isolate database
- *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}/isolates/{isolate_id}/allele_designations/{locus} Retrieve full allele designation record
- *GET /db/{database}/isolates/{isolate_id}/allele_ids* Retrieve allele identifiers
- *GET /db/{database}/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}/isolates/{isolate_id}/history Retrieve isolate update history
- GET /db/{database}/contigs/{contig_id} Retrieve contig record
- GET /db/{database}/fields Retrieve list of isolate provenance field descriptions
- GET /db/{database}/fields/{field} Retrieve values set for a provenance field
- GET /db/{database}/users/{user_id} Retrieve user information
- *GET /db/{database}/curators* Retrieve list of curators of the database
- 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

17.3.1 GET / or /db - List site resources

Required route parameters: None

Optional query parameters: None

Example request URI: https://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

17.3.2 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: https://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
- genomes [string] URI to genome records
- schemes [string] URI to list of schemes
- loci [string] URI to list of loci
- projects [string] URI to list of projects

17.3.3 GET /db/{database}/classification_schemes - List classification schemes

Required route parameter: database [string] - Database configuration name

Optional parameters: None

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/classification_schemes

Response: Object containing:

- records [integer] Number of classification schemes.
- classification_schemes [array] List of URIs to classificaton schemes.

17.3.4 GET /db/{database}/classification_schemes/{classification_scheme_id} - Retrieve classification scheme information and groups

Sequence definition databases only.

Required route parameters:

- database [string] Database configuration name
- · classification_scheme_id [integer] Classification scheme id number

Optional parameters: None

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/classification_schemes/1

Response: Object containing some or all of:

• id [integer] - Classification scheme id

- name [text] Name of classification scheme
- description [text] Description of classification scheme
- relative_threshold [boolean] True if a relative thresold is used
- inclusion_threshold [integer] The threshold for number of loci difference used to group
- groups [array] (sequence definition databases only) list of group objects consisting of:
 - id [integer] group id
 - profiles [array] list of URIs to profiles belonging to the group

17.3.5 GET /db/{database}/classification_schemes/{classification_scheme_id}/groups/{group_id - List isolates belonging to a classification scheme group

Isolate databases only.

Required route parameters:

- database [string] Database configuration name
- · classification_scheme_id [integer] Classification scheme id number
- group_id [integer] Group id number

Optional parameters:

- page [integer]
- page_size [integer]
- return_all [integer] Set to non-zero value to disable paging.

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/classification_schemes/4/groups/65

Response: Object containing of:

- records [integer] 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)

17.3.6 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.
- alleles_added_after [date] Include only loci with alleles added after specified date (ISO 8601 format). Only recognized in sequence definition databases.
- alleles_updated_after [date] Include only loci with alleles last modified after specified date (ISO 8601 format). Only recognized in sequence definition databases.

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci

Response: Object containing:

- records [integer] 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)

Note: See also the *scheme specific version*, allowing filtering by date of last allele update for just the loci that are members of a scheme.

17.3.7 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: https://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)

17.3.8 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.
- added_after [date] Include only alleles added after specified date (ISO 8601 format).
- added_on [date] Include only alleles added on specified date (ISO 8601 format).
- updated_after [date] Include only alleles last modified after specified date (ISO 8601 format).

• updated_on [date] - Include only alleles last modified on specified date (ISO 8601 format).

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci/abcZ/alleles

Response: Object containing:

- records [integer] Number of alleles.
- last_updated [date] Latest allele addition/modification date (ISO 8601 format).
- 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)

17.3.9 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:

- added_after [date] Include only alleles added after specified date (ISO 8601 format).
- added_on date] Include only alleles added on specified date (ISO 8601 format).
- updated_after [date] Include only alleles last modified after specified date (ISO 8601 format).
- updated_on [date] Include only alleles last modified on specified date (ISO 8601 format).

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci/abcZ/alleles_fasta

Response: FASTA format file of allele sequences

17.3.10 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: https://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)

17.3.11 POST /db/{database}/loci/{locus}/sequence - Query sequence to identify allele

Required route parameters:

- database [string] Database configuration name
- locus [string] Locus name

Required additional parameters (JSON-encoded in POST body):

• sequence [string] - Sequence string or base64-encoded FASTA file

Optional parameters (JSON-encoded in POST body):

- · details [true/false] Return detailed exact match parameters
- base64 [true/false] Sequence is a base64-encoded FASTA file

Response: Object containing the following key/value pairs:

- exact_matches [array] list of match objects, each consisting of:
 - allele_id
 - href URI to allele record.

additionally if 'details' parameter passed:

- start start position on query
- end end position on query
- orientation forward/reverse
- length length of matched allele
- contig contig name if FASTA file is uploaded

If the locus is linked to field data in client isolate databases, there may also be an object called 'linked_data' containing values and frequencies of the field for the returned allele.

- best_match [object] consisting of key/value pairs (if no exact matches)
 - allele_id
 - href URI to allele record.
 - start start position on query (predicted taking account of allele length)
 - end end position on query (predicted taking account of allele length)

- orientation forward/reverse
- length length of matched allele
- alignment length of BLAST alignment
- mismatches number of mismatches
- identity %identity of match
- gaps number of gaps in alignment

17.3.12 POST /db/{database}/sequence - Query sequence to identify allele without specifying locus

Required route parameters:

• database [string] - Database configuration name

Required additional parameters (JSON-encoded in POST body):

• sequence [string] - Sequence string or base64-encoded FASTA file

Optional parameters (JSON-encoded in POST body):

- details [true/false] Return detailed exact match parameters
- base64 [true/false] Sequence is a base64-encoded FASTA file

Response:

- exact_matches [object] consisting of locus keys, each consisting of array of match objects consisting of:
 - allele_id
 - href URI to allele record.

additionally if 'details' parameter passed:

- start start position on query
- end end position on query
- orientation forward/reverse
- length length of matched allele
- contig contig name if FASTA file is uploaded

If the locus is linked to field data in client isolate databases, there may also be an object called 'linked_data' containing values and frequencies of the field for the returned allele.

Note: This method only supports exact matches. If no match is indicated for a specific locus, use the *locus-specific call* to identify the closest match.

17.3.13 GET /db/{database}/sequences - Get summary of defined sequences

Required route parameter: database [string] - Database configuration name

Optional parameters:

• added_after [date] - Count only alleles added after specified date (ISO 8601 format).

- added_on [date] Count only alleles added on specified date (ISO 8601 format).
- updated_after [date] Count only alleles last modified after specified date (ISO 8601 format).
- updated_on [date] Count only allele updated on specified date (ISO 8601 format).

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/sequences

Response: Object containing a subset of the following key/value pairs:

- loci [string] URI to list of loci
- records [integer] Number of alleles defined
- last_updated [date] Latest allele addition/modification date (ISO 8601 format).

17.3.14 GET /db/{database}/schemes - List schemes

Required route parameter: database [string] - Database configuration name

Optional parameters:

• with_pk [integer] - Set to non-zero value to only show indexed schemes, i.e. those with a primary key field that defines each unique combination of alleles, e.g. MLST.

Example request URI: https://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]

17.3.15 GET /db/{database}/schemes/{scheme_id} - Retrieve scheme information

Includes links to allelic profiles (in sequer databases, if appropriate). Required route parameters:

- database [string] Database configuration name
- scheme_id [integer] Scheme id number

Optional parameters:

- added_after [date] Count only profiles added after specified date (ISO 8601 format).
- added_on [date] Count only profiles added on specified date (ISO 8601 format).
- updated_after [date] Count only profiles last modified after specified date (ISO 8601 format).
- updated_on [date] Count only profiles updated on specified date (ISO 8601 format).

Example request URI: https://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
- records [integer] Number of profiles
- last_added [date] Latest profile addition/modification date (ISO 8601 format).
- last_updated [date] Latest profile addition/modification date (ISO 8601 format).

17.3.16 GET /db/{database}/schemes/{scheme_id}/loci - Retrieve scheme loci

Required route parameters:

- database [string] Database configuration name
- scheme_id [integer] Scheme id number

Optional parameters:

- alleles_added_after [date] Include only loci with alleles added after specified date (ISO 8601 format). Only recognized in sequence definition databases.
- alleles_updated_after [date] Include only loci with alleles last modified after specified date (ISO 8601 format). Only recognized in sequence definition databases.

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/loci

Response: Object containing:

- records [integer] Number of loci
- loci [array] List of URIs to defined locus records.

17.3.17 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: https://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

17.3.18 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.
- added_after [date] Include only profiles added after specified date (ISO 8601 format).
- added_on [date] Include only profiles added on specified date (ISO 8601 format).
- updated_after [date] Include only profiles last modified after specified date (ISO 8601 format).
- updated_on [date] Include only profiles last modified on specified date (ISO 8601 format).

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/profiles

Response: Object containing:

- records [integer] Number of profiles
- last_updated [date] Latest profile addition/modification date (ISO 8601 format).
- 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)

Note: This method also supports content negotiation. If the request accepts header includes TSV or CSV, then the call is redirected to /*db*/{*database*}/*scheme_id*}/*profiles_csv*.

17.3.19 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:

• added_after [date] - Include only profiles added after specified date (ISO 8601 format).

- added_on [date] Include only profiles added on specified date (ISO 8601 format).
- updated_after [date] Include only profiles last modified after specified date (ISO 8601 format).
- updated_on [date] Include only profiles last modified on specified date (ISO 8601 format).

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/profiles_csv

Response: Tab-delimited text file of allelic profiles

17.3.20 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: https://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)

17.3.21 POST /db/{database}/schemes/{scheme_id}/sequence - Query sequence to extract allele designations/fields for a scheme

Required route parameters:

- database [string] Database configuration name
- scheme_id [integer] Scheme id

Required additional parameters (JSON-encoded in POST body):

• sequence [string] - Sequence string or base64-encoded FASTA file

Optional parameters (JSON-encoded in POST body):

- details [true/false] Return detailed exact match parameters
- base64 [true/false] Sequence is a base64-encoded FASTA file

Response: Object containing the following key/value pairs:

• exact_matches [array] - list of match objects, each consisting of:

- allele_id
- href URI to allele record.

additionally if 'details' parameter passed:

- start start position on query
- end end position on query
- orientation forward/reverse
- length length of matched allele
- contig contig name if FASTA file is uploaded

If the locus is linked to field data in client isolate databases, there may also be an object called 'linked_data' containing values and frequencies of the field for the returned allele.

Example curl call to upload a FASTA file 'contigs.fasta' and extract MLST results from Neisseria database:

```
(echo -n '{"base64":true,"sequence": "'; base64 contigs.fasta; echo '"}') |
curl -s -H "Content-Type: application/json" -X POST "https://rest.pubmlst.org/db/
→pubmlst_neisseria_seqdef/schemes/1/sequence" -d @-
```

Note: This method only supports exact matches. If no match is indicated for a specific locus, use the *locus-specific call* to identify the closest match.

17.3.22 POST /db/{database}/schemes/{scheme_id}/designations - Query allelic profile to extract fields for a scheme

Required route parameters:

- database [string] Database configuration name
- scheme_id [integer] Scheme id

Required additional parameters (JSON-encoded in POST body):

- · designations [object] consisting of
 - locus objects each containing an array of alleles (see example)

Response: Object containing the following key/value pairs:

- exact_matches [object] consisting of locus values, each consisting of an array of allele values:
 - allele_id [string]

If the locus is linked to field data in client isolate databases, there may also be an object called 'linked_data' containing values and frequencies of the field for the returned allele.

• fields [object] - consisting of key/value pairs of scheme fields (if defined)

Example curl call to query an allelic profile and extract MLST results from Neisseria database:

```
curl -s -H "Content-Type: application/json" -X POST "https://rest.pubmlst.org/db/

→pubmlst_neisseria_seqdef/schemes/1/designations" -d '{"designations":{"abcZ":[{

→"allele":"2"}],"adk":[{"allele":"3"}],"aroE":[{"allele":"4"}],"fumC":[{"allele":"3"}],"gdh":[{"allele":"8"}],"pdhC":[{"allele":"4"}],"pgm":[{"allele":"6"}]}}'
```

17.3.23 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.
- added_after [date] Include only isolates added after specified date (ISO 8601 format).
- added_on [date] Include only isolates added on specified date (ISO 8601 format).
- include_old_versions [integer] Set to 1 to include old record versions (the default is to only include new versions)
- updated_after [date] Include only isolates last modified after specified date (ISO 8601 format).
- updated_on [date] Include only isolates updated on specified date (ISO 8601 format).

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates

Response: Object containing:

- records [integer] 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)

17.3.24 GET /db/{database}/isolates/{isolate_id} - Retrieve isolate record

Required route parameters:

- database [string] Database configuration name
- isolate_id [integer] Isolate identifier

Optional parameter:

• provenance_only [integer] - Set to non-zero value to only return provenance metadata

Example request URI: https://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* 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 some of 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
 - classification_schemes [object] consisting of key/value pairs, where each key is the name of the classification scheme and the value is an object consisting of:
 - * href [string] URI to classification scheme description
 - * groups [array] list of group objects consisting of:
 - · group [integer] group id
 - · records [integer] number of isolates in group
 - · isolates [string] URI to classification group record containing URIs to member isolate records
- projects [array] list of project objects, each containing the following:
 - id [string] URI to project information
 - description [string] description of project
- history [string] URI to isolate history record
- new_version [string] URI to newer version of record
- old_version [string] URI to older version of record

17.3.25 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: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/allele_designations

Response: Object containing:

- records [integer] 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)

17.3.26 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: https://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)

17.3.27 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: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/allele_ids

Response: Object containing:

- records [integer] 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)

17.3.28 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: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/schemes/1/allele_designations

Response:

- records [integer] Number of allele designation objects
- allele_designations [array] List of *allele designation objects* for each locus in the specified scheme that has been designated.

17.3.29 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: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/schemes/1/allele_ids

Response:

- records [integer] 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.

17.3.30 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: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/contigs

Response: Object containing:

- records [integer] 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)

17.3.31 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: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/contigs_fasta?header= original_designation

Response: FASTA format file of isolate contig sequences

17.3.32 GET /db/{database}/isolates/{isolate_id}/history - Retrieve isolate update history

Required route parameters:

- database [string] Database configuration name
- isolate_id [integer] Isolate identifier

Optional parameters: None

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/history

Response: Object containing:

- records [integer] Number of updayes
- contigs [array] List of update objects each consisting of the following key/value pairs:
 - curator [string] URI to user details of curator
 - timestamp [string] Time of update
 - actions [array] List of update descriptions [strings]

17.3.33 GET /db/{database}/genomes - Retrieve list of isolate records that have genome assemblies

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.
- added_after [date] Include only isolates added after specified date (ISO 8601 format).
- added_on [date] Include only isolates added on specified date (ISO 8601 format).
- include_old_versions [integer] Set to 1 to include old record versions (the default is to only include new versions)

- updated_after [date] Include only isolates last modified after specified date (ISO 8601 format).
- updated_on [date] Include only isolates updated on specified date (ISO 8601 format).
- genome_size [integer] Filter to only include records with a sequence bin of at least the specified size (default is 500,000bp).

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/genomes

Response: Object containing:

- records [integer] 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)

17.3.34 POST /db/{database}/isolates/search - Search isolate database

Required route parameters:

• database [string] - Database configuration name

Optional parameters (appended to URI):

- page [integer]
- page_size [integer]
- return_all [integer] Set to non-zero value to disable paging.

Query parameters (JSON-encoded in POST body):

You must include at least one query parameter.

Parameter names in the following forms are supported:

- field.{field} key/value pairs for provenance fields. Supported field names can be found by calling the */fields route*. The fields will vary by database.
- locus.{locus} key/value pairs of locus and its allele designation. Supported locus names can be found by calling the */loci route*.
- scheme_id}.{scheme_field} key/value pairs of scheme fields and their values. Supported field names can be determined by following routes from the */schemes route*.

Example method call using curl: The following searches for *Neisseria* ST-11 isolates from Europe in 2015 (MLST is scheme#1 in this database).

```
curl -s -H "Content-Type: application/json" -X POST \
"https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/search" \
-d '{"field.continent":"europe","field.year":2015,"scheme.1.ST":11}'
```

Response: Object containing:

- records [integer] 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)

17.3.35 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: https://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)
- loci [array] list of sequence tag objects consisting of:
 - locus [string] URI to locus description
 - locus_name [string]
 - start [integer]
 - end [integer]
 - direction [string] forward/reverse
 - complete [boolean] true/false

17.3.36 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: https://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
- values [string] URI to list of used field values

17.3.37 GET /db/{database}/fields/{field} - Retrieve values set for a provenance field

Required route parameters:

- database [string] Database configuration name
- field [string] Provenance metadata field name

Optional parameters:

- page [integer]
- page_size [integer]
- return_all [integer] Set to non-zero value to disable paging.

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/fields/country

Response: Object containing:

- records [integer] Number of values
- values [array] List of values used in 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)

17.3.38 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: https://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 (may be hidden depending on server configuration)

17.3.39 GET /db/{database}/curators - Retrieve list of curators

Required route parameters:

• database [string] - Database configuration name

Optional parameters: None

Example request URI: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/curators

Response: Object containing:

- records [integer] Number of curators
- curators [array] List of URIs to user records.

17.3.40 GET /db/{database}/projects - Retrieve list of projects

Required route parameter: database [string] - Database configuration name

Optional parameters: None

Example request URI: https://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

17.3.41 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: https://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.

17.3.42 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: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/projects/3/isolates

Response: Object containing:

- records [integer] 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)

17.3.43 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: https://rest.pubmlst.org/db/pubmlst_neisseria_isolates/submissions

Response: Object containing:

- records [integer] 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)

17.3.44 POST /db/{database}/submissions - create new submission

Required route parameter: database [string] - Database configuration name

Required additional parameters (JSON-encoded in POST body):

- type [string] either:
 - alleles (sequence definition databases only)
 - profiles (sequence definition databases only)
 - isolates (isolate databases only)
 - genomes (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 [integer] - 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

Genome submissions

• isolates [string] - tab-delimited isolate data - this should include a header line containing each field or locus included as well as for 'assembly_filename' and 'sequence_method'. The 'sequence_method' should be either '454', 'Illumina', 'Ion Torrent', 'PacBio', 'Oxford Nanopore', 'Sanger', 'Solexa', 'SOLiD', or 'other'. Follow-ing submission, contig files should be uploaded with the same names as set for 'assembly_filename'. This can be done using the *file upload route*.

Optional parameters:

- message [string] correspondence to the curator
- email [integer] 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

For genome submissions, the response object will also contain:

- missing_files [array] List of filenames that need to be uploaded to complete the submission. These filenames are defined in the 'assembly_filename' field of the isolate record upload. The files should contain the contig assemblies.
- message [string] 'Please upload missing contig files to complete submission.'

17.3.45 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: https://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.

17.3.46 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:	https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/submissions/BIGSdb_
20151013081	836_14559_	14740	

Response: message [string] - 'Submission deleted.'

17.3.47 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: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/submissions/BIGSdb_20151013081836_14559_14740/messages

Response: Array of correspondence objects in time order. Each contains:

- user [string] URI to user details of user
- timestamp [string]
- message [string]

17.3.48 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 (JSON-encoded in POST body):

• message [string] - Message text

Optional parameters: None

Response: message [string] - 'Message added.'

17.3.49 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: https://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/submissions/BIGSdb_ 20151013081836_14559_14740/files

Response: Array of URIs to files

17.3.50 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 (JSON-encoded in POST body):

- filename [string] Name of file to store within submission
- upload [base64 encoded data] Raw file data

Optional parameters: None

Response: message [string] - 'File uploaded.'

17.3.51 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

17.3.52 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.'

17.4 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.

Note: There are Python and Perl example scripts available at https://github.com/kjolley/BIGSdb/tree/develop/scripts/ rest_examples to demonstrate and test OAuth authentication.

17.4.1 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

17.4.2 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.

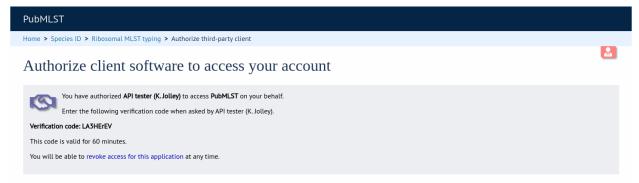
17.4.3 Getting user authorization

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=fKFm0WNhCfbEX8zQm6qhDA8K23FOWDGE

The user will be asked if they wish to grant access to the application on their behalf:

PubMLST			
Home > Species ID > Ribosomal MLST typing > Authorize third-party client			
Authorize client software to access your account	.		
Do you wish for the following application to access data on your behalf? Application Resource API tester (K. Jolley) PubMLST 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.

17.4.4 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 either 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

17.4.5 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

17.4.6 Accessing protected resources

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')

CHAPTER 18

Frequently asked questions (FAQs)

18.1 General

1. What is the minimum specification of hardware required to run BIGSdb?

The software will run on fairly modest hardware. 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 multiple cores (depending on the settings in bigsdb.conf), 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* is usually 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: 40 cores, 128GB RAM
- database server: 80 cores, 1TB RAM, 7TB ZFS RAID-Z2 NVMe local storage

2. Why might icons be missing when using Internet Explorer?

This can occur if you have Compatibility Mode enabled. BIGSdb generates valid HTML5 and Compatibility Mode should not be used. Please ensure this is not enabled in the Internet Explorer tools section.

18.2 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.

18.3 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 and private records.

CHAPTER 19

Appendix

19.1 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.
- >=
 - Greater than or equal the search term.

• <

- Less than the search term.
- <=
 - Less than or equal 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).

19.2 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:

- alternative start codon
 - A start codon other than ATG, GTG, or TTG is used. This can be the case with some yeasts.
- · 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.
- indel
 - Insertion/deletion in sequence that is uncommon compared to other alleles.
- 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.
- · no stop codon
 - No stop codon in immediate vicinity of usual stop.
- 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).

19.3 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:

- alternative start codon
 - A start codon other than ATG, GTG, or TTG is used. This can be the case with some yeasts.
- 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.
- indel
 - Insertion/deletion in sequence that is uncommon compared to other alleles.
- 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.
- · no stop codon
 - No stop codon in immediate vicinity of usual stop.
- · 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).

CHAPTER 20

Database schema

- Sequence definition database
- Isolate database

Index

A

access control lists, 51 restricting, 51 adding classification groups, 86 isolates, 134 locus, 59, 61, 62, 67, 91 MLST scheme, 84 schemes, 70 allele definition records, 209 allele designations count, 246 query, 245 status, 248 allele sequence identify, 215 allele sequences adding, 112 alleles, 3 list query, 225 retiring, 121 un-retiring, 122 allelic profiles isolates, 257 annotation status query, 249 API authentication access token, 435 accessing protected resources, 437 consumer key, 434 request token, 434 session token, 436 user authorization, 435 API resources add submission correspondence, 432 create new submission, 429 DELETE /db/{database}/submissions/{submissi ∂a_{1}^{1} id},

DELETE /db/{database}/submissions/{submission_i 433 delete submission record, 431 delete submission supporting file, 433 download alleles in FASTA format, 410 download allelic profiles in CSV (tab-delimited) format, 415 download contigs in FASTA format, 422 download submission supporting file,433 GET /,405 GET /db, 405 GET /db/{database},406 GET /db/{database}/classification_schemes, 406 GET /db/{database}/classification_schemes/{clas 406 GET /db/{database}/classification_schemes/{clas 407 GET /db/{database}/contigs/{contig_id}, 425 GET /db/{database}/curators, 427 GET /db/{database}/fields, 425 GET /db/{database}/fields/{field}, 426 GET /db/{database}/genomes, 423 GET /db/{database}/isolates,417 GET /db/{database}/isolates/{isolate_id}, 418 GET /db/{database}/isolates/{isolate_id}/allele 419 GET /db/{database}/isolates/{isolate_id}/allele 420GET /db/{database}/isolates/{isolate_id}/allele GET /db/{database}/isolates/{isolate_id}/contig

431

422 list schemes, 413 GET /db/{database}/isolates/{isolate_id}//contige_factores,405 422 POST /db/{database}/isolates/search, GET /db/{database}/isolates/{isolate_id}/history, 423 POST /db/{database}/loci/{locus}/sequence, GET /db/{database}/isolates/{isolate id}/scheme id}/allele designations, POST /db/{database}/schemes/{scheme id}/designa 421 GET /db/{database}/isolates/{isolate_id}/scheme_id}/allele_ids, 421 POST /db/{database}/schemes/{scheme id}/sequenc GET /db/{database}/loci,407 416 GET /db/{database}/loci/{locus},408 POST /db/{database}/sequence, 412 GET /db/{database}/loci/{locus}/alleles, POST /db/{database}/submissions, 429 409 POST /db/{database}/submissions/{submission_id} GET /db/{database}/loci/{locus}/alleles/{allale_id}, POST /db/{database}/submissions/{submission_id} 410 GET /db/{database}/loci/{locus}/alleles_fasta2 410 query allele sequence, 411 GET /db/{database}/projects,427 query allele sequence without GET /db/{database}/projects/{project_id}, specifying locus, 412 427 query scheme designations, 417 GET /db/{database}/projects/{project_id}dieoyaseheme sequences,416 retrieve allele identifiers, 421 428 GET /db/{database}/schemes, 413 retrieve classification scheme GET /db/{database}/schemes/{scheme id}, information and groups, 406, 407 413 retrieve contig record, 425 GET /db/{database}/schemes/{scheme_id}/fiefdievield}, allele designation record, 420 414 GET /db/{database}/schemes/{scheme_id}/loetrieve full allele information, 414 410 GET /db/{database}/schemes/{scheme_id}/prefilese information about scheme 414 field, 414 GET /db/{database}/schemes/{scheme_id}/prefiles#{psofife_id&prd,418 416 retrieve isolate update history, 423 GET /db/{database}/schemes/{scheme_id}/prefilesecsist of allele designations, 419 415 retrieve list of alleles defined GET /db/{database}/sequences,412 GET /db/{database}/submissions, 428 for a locus, 409 GET /db/{database}/submissions/{submissionetideve list of contigs,422 430 retrieve list of curators, 427 GET /db/{database}/submissions/{submissionetrdev€ilest of isolate 432 provenance field descriptions, GET /db/{database}/submissions/{submission_#d5/files/{filename}, 433 retrieve list of isolate records, GET /db/{database}/submissions/{submission_4d7/messages, 432 retrieve list of isolate records GET /db/{database}/users/{user_id}, that have genome assemblies, 423 427 retrieve list of isolates get summary of defined sequences, belonging to a project, 428 412 retrieve list of projects, 427 list allelic profiles defined for retrieve list of scheme allele identifiers, 421 scheme, 414 list classification schemes, 406 retrieve list of submissions, 428 list database resources, 406 retrieve list of supporting files list loci, 407 uploaded for submission, 432

retrieve locus record, 408 retrieve project information, 427 retrieve scheme allele designation records. 421 retrieve scheme information, 413 retrieve scheme loci, 414 retrieve specific allelic profile record, 416 retrieve submission correspondence, 432 retrieve submission record, 430 retrieve user information, 427 retrieve values set for a provenance field, 426 search isolate database, 424 upload submission supporting file, 432 auto allele definer, 192 stop, 195 automated assignment scheme profiles, 85 autotagger, 189 stop, 195

B

batch profile definitions query, 235 batch uploading contigs - multiple isolates, 152 BLAST, 291 BLAST caches refreshing, 108 bookmarks, 260 breakdown provenance field, 304 sequence bin, 360 two-field, 363 browse isolates, 241 scheme profiles, 226 BURST. 297

С

caching schemes, 54 classification groups, 4 adding, 86 client authorization RESTful interface, 107 client databases, 79 clustering core genome, 89 codon usage, 301 Locus Explorer, 345 composite fields,98
config-specific
 downloads,109
configuration
 export,105
configuration settings
 validation,104
core genome
 clustering,89
count
 allele designations,246
 sequence tags,251

D

defining exemplar alleles, 191 deleting isolates, 139, 142 profiles, 130 downloads config-specific, 109

Ε

exemplar alleles defining, 191 export configuration, 105 extended attributes locus, 68 provenance fields, 100

F

filters,256

G

gene presence;,315
Genome Comparator,321
genome filtering,92
 in silico hybridization,95
 in silico PCR,92
GrapeTree
 minimum-spanning trees,330
groups
 scheme,76

Η

hosts mapping, 54

I

identify
 allele sequence, 215
 sequence type, 217
in silico hybridization, 95

in silico PCR, 92 in silico PCR analysis, 334 isolate records, 205 isolate aliases, 146 isolate record options, 270 isolates adding, 134 allelic profiles, 257 browse, 241 deleting, 139, 142 query, 242 retiring, 143 un-retiring, 144 updating, 138, 141 uploading contigs, 149 iTOL phylogenetic trees, 338

K

kiosk mode,41

L

linking remote contigs, 155
list query
 alleles, 225
loci, 3
 modifying display, 274
locus
 adding, 59, 61, 62, 67, 91
 adding; copying existing record, 66
 extended attributes, 68
Locus Explorer, 342
 codon usage, 345
 polymorphic sites, 342
 translated sequences, 346
locus positions
 setting, 96

Μ

main results table options, 268 mapping hosts, 54 Microreact spatio-temporaral trees, 347 minimum-spanning trees GrapeTree, 330 PhyloViz online, 350 MLST, 4 MLST scheme adding, 84 mod_perl, 54 modifying display loci,274 schemes,274

0

offline curation auto allele definer, 192 autotagger, 189 options, 266 isolate record, 270 main results table, 268 provenance fields, 271 query, 273

Ρ

partitioning sets,56 passwords setting, 52 setting; first user, 52 performance caching schemes, 54 mod_perl, 54 permissions, 47 locus curation, 50 scheme curation, 50phylogenetic trees iTOL, 338 PhyloViz online minimum-spanning trees, 350 plugins enabling, 52 polymorphic sites Locus Explorer, 342 polymorphisms, 353 profile records, 211 profiles,4 deleting, 130 retiring, 131 un-retiring, 132 updating, 129 projects, 160 provenance field breakdown. 304 provenance fields extended attributes, 100 options, 271 publications, 147, 263

Q

query allele designations,245 batch profile definitions,235 isolates, 242
options, 273
scheme profiles, 228
sequence tags, 253
ST definitions from allelic
 profiles, 232

R

```
records
   allele definition, 209
   isolate, 205
   profile, 211
   sequence bin, 212
   sequence tag, 210
refreshing
   BLAST caches, 108
RESTful interface
   client authorization, 107
retiring
   alleles, 121
   isolates, 143
   profiles, 131
rMLST
   species identification, 356
```

S

scheme groups, 76 scheme profiles automated assignment, 85 browse, 226 query, 228 schemes,4 adding,70 caching, 54 modifying display, 274 sequence bin breakdown, 360 query, 250 records, 212 sequence similarity determining, 237 sequence tag records, 210 sequence tags, 4 count, 251 query, 253 sequence type identify, 217 sets.5 partitioning, 56 sparsely-populated fields, 39 spatio-temporaral trees Microreact, 347

species identification
 rMLST, 356
ST definitions from allelic profiles
 query, 232
stop
 auto allele definer, 195
 autotagger, 195

Т

translated sequences Locus Explorer, 346 two-field breakdown, 363

U

```
un-retiring
   alleles, 122
   isolates, 144
   profiles, 132
unique combinations, 366
updates
   disabling, 53
updating
   isolates, 138, 141
   profiles, 129
uploading contigs
   isolates, 149
user groups, 47
user projects, 278
user types, 47
users
   adding, 111
```