# **BIGSdb Documentation**

Release 1.38.0

**Keith Jolley** 

Dec 15, 2022

# CONTENTS

1	Conc	epts and terms 3
	1.1	BIGSdb
	1.2	Loci
	1.3	Alleles
	1.4	Schemes
	1.5	Profiles
	1.6	Classification groups
	1.7	Sequence tags
	1.8	Sets
2	BIGS	Sdb dependencies 7
	2.1	Required packages
3	Insta	llation and configuration of BIGSdb 9
	3.1	Software installation
	3.2	Configuring PostgreSQL
	3.3	Setting global connection parameters
	3.4	Site-specific configuration
	3.5	Setting up the offline job manager
	3.6	Setting up the submission system
	3.7	Setting up a site-wide user database
	3.8	Periodically delete temporary files
	3.9	Prevent preference database getting too large 16
	3.10	Log file rotation
	3.11	Upgrading BIGSdb
	3.12	Running the BIGSdb RESTful interface    17
4	Data	base setup 21
	4.1	Creating databases
	4.2	Database-specific configuration
	4.3	XML configuration attributes used in config.xml
	4.4	Over-riding global defaults set in bigsdb.conf
	4.5	Over-riding values set in config.xml
	4.6	Setting field validation rules
	4.7	Sparsely-populated fields
	4.8	Kiosk mode
	4.9	User authentication
	4.10	Setting up the admin user
	4.11	Retrieving PubMed citations from NCBI
	4.12	Configuring access to remote contigs

	4.13	Setting up front-end and query dashboards
5		inistrator's guide 57
	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       89
	5.19	Workflow for setting up a MLST scheme    94
	5.20	Automated assignment of scheme profiles
	5.20	Scheme profile clustering - setting up classification schemes
	5.22	Defining new loci based on annotated reference genome
	5.22	Setting up LINcode definitions for cgMLST schemes
	5.23	Genome filtering
	5.24	Setting locus genome positions
	5.25	Defining composite fields
	5.20	Extended provenance attributes (lookup tables)
	5.28	Sequence bin attributes
	5.29	Checking external database configuration settings
	5.30	Exporting table configurations
	5.31	Authorizing third-party client software to access authenticated resources
	5.32	BLAST caches
	5.33	Config-specific file downloads
6	Cura	tor's guide 125
U	6.1	Adding new sender details
	6.2	Adding new allele sequence definitions
	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       141         Updating and deleting scheme profile definitions       144
	6.8 6.9	
		0 · · · · · · · · · · · · · · · · · · ·
	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
	6.16	Un-retiring isolate identifiers

		Setting alternative names for isolates (aliases)	
	6.18	Linking isolate records to publications	
		Uploading sequence contigs linked to an isolate record	
	6.20	Batch uploading sequence contigs linked to multiple isolate records	
	6.21	Linking remote contigs to isolate records	
	6.22	Automated web-based sequence tagging	
	6.23	Projects	
	6.24	Isolate record versioning	
	6.25	Populating geographic coordinate lookup values	181
7	Cura	ting data submitted via the automated submission system	183
	7.1	Alleles	
	7.2	Profiles	
	7.3	Isolates	198
8			203
	8.1	Automated offline sequence tagging	
	8.2	Defining exemplar alleles	
	8.3		207
	8.4		209
	8.5	Predicting species based on rMLST analysis	
	8.6	Cleanly interrupting offline curation	
	8.7	Uploading contigs from the command line	
	8.8	Populating geographic point lookup values	213
9	Defin	ition downloads	215
-	9.1	Allele sequence definitions	
	9.2	Scheme profile definitions	
		*	
10			221
10	10.1	Isolate records	221
10	10.1 10.2	Isolate records	221 225
10	10.1 10.2 10.3	Isolate records       .         Allele definition records       .         Sequence tag records       .	221 225 227
10	10.1 10.2 10.3 10.4	Isolate records	221 225 227 228
10	10.1 10.2 10.3 10.4	Isolate records	221 225 227
	10.1 10.2 10.3 10.4 10.5	Isolate records	221 225 227 228 229
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b>	Isolate records       2         Allele definition records       2         Sequence tag records       2         Profile records       2         Sequence bin records       2         ying data       2	221 225 227 228 229 <b>231</b>
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1	Isolate records	221 225 227 228 229 <b>231</b> 231
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1	Isolate records       2         Allele definition records       2         Sequence tag records       2         Profile records       2         Sequence bin records       2         ying data       2	221 225 227 228 229 <b>231</b> 231 234
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2	Isolate records	221 225 227 228 229 <b>231</b> 231 234 236
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3	Isolate records	221 225 227 228 229 <b>231</b> 231 234 236
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4	Isolate records	221 225 227 228 229 <b>231</b> 231 234 236 243
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5	Isolate records	221 225 227 228 229 <b>231</b> 231 234 236 243 244
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6	Isolate records	221 225 227 228 229 <b>231</b> 234 236 243 244 248
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9	Isolate records	221 225 227 228 229 <b>231</b> 234 236 243 244 248 252 254 252
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 11.10	Isolate records	221 225 227 228 229 <b>231</b> 234 236 243 244 252 254 252 254 257 259
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 11.10 11.11	Isolate records       Allele definition records         Sequence tag records       Profile records         Profile records       Sequence bin records         Sequence bin records       Sequence bin records         ying data       Querying sequences to determine allele identity         Querying multiple sequences to identify allele identities       Searching for specific allele definitions         Browsing scheme profile definitions       Guerying scheme profile definitions         Identifying allelic profile definitions       Identifying allele differences         Browsing isolate data       Querying isolate data         Bookmarking an isolate query       Searching an isolate query	221 225 227 228 229 <b>231</b> 234 236 243 244 252 254 252 254 257 259 276
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 11.10 11.11 11.12	Isolate records       Allele definition records         Sequence tag records       Profile records         Profile records       Sequence bin records         Sequence bin records       Sequence bin records         ying data       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       Identifying allele differences         Browsing isolate data       Querying isolate data         Bookmarking an isolate query       Retrieving isolates by linked publication	221 225 227 228 229 <b>231</b> 234 236 243 244 252 254 252 254 257 259 276 279
	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 11.10 11.11 11.12	Isolate records       Allele definition records         Sequence tag records       Profile records         Profile records       Sequence bin records         Sequence bin records       Sequence bin records         ying data       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       Identifying allele differences         Browsing isolate data       Querying isolate data         Bookmarking an isolate query       Retrieving isolates by linked publication	221 225 227 228 229 <b>231</b> 234 236 243 244 252 254 252 254 257 259 276
11	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 11.10 11.11 11.12 11.13	Isolate records       Allele definition records         Sequence tag records       Profile records         Profile records       Sequence bin records         Sequence bin records       Sequence bin records         ying data       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       Identifying allele differences         Batch profile queries       Investigating allele differences         Browsing isolate data       Querying isolate data         Bookmarking an isolate query       Retrieving isolates by linked publication         User-configurable options       User-configurable options	221 225 227 228 229 <b>231</b> 234 236 243 244 252 254 259 259 276 279 282
11	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 11.10 11.11 11.12 11.13 <b>User</b>	Isolate records	221 225 227 228 229 231 234 236 243 244 248 254 254 254 257 259 276 279 282 279 282
11	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 11.10 11.11 11.12 11.13 <b>User</b> 12.1	Isolate records       Allele definition records         Sequence tag records       Profile records         Profile records       Sequence bin records         Sequence bin records       Sequence bin records         ying data       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       Identifying allele differences         Browsing isolate data       Querying isolate data         Querying isolate data       Power sequences         Browsing isolate data       Searching an isolate query         Retrieving isolates by linked publication       User-configurable options	221 225 227 228 229 231 234 236 243 244 252 254 252 254 257 259 276 279 282 279 282 295
11	10.1 10.2 10.3 10.4 10.5 <b>Quer</b> 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 11.10 11.11 11.12 11.13 <b>User</b>	Isolate records       Allele definition records         Sequence tag records       Profile records         Profile records       Sequence bin records         Sequence bin records       Sequence bin records         ying data       Querying sequences to determine allele identity         Querying multiple sequences to identify allele identities       Searching for specific allele definitions         Browsing scheme profile definitions       Guerying scheme profile definitions         Identifying allelic profile definitions       Identifying allele identifies         Batch profile queries       Investigating allele differences         Browsing isolate data       Ouerying isolate data         Querying isolate data       Ouerying isolate data         Querying isolate by linked publication       User-configurable options         Projects       Adding isolates to a user project following a query	221 225 227 228 229 231 234 234 243 244 252 254 257 259 276 279 282 296 297

		Allowing other users to share your project    2      Deleting a user project    3	
13	Priva	te records 3	03
	13.1	Uploading private records	03
		Modifying private records	
	13.3	Sharing access to private records	06
14	Data	analysis plugins 3	07
		BLAST	
		BURST	
		Codon usage	
		Field breakdown    32	
		Gene Presence	
		Genome comparator	
	14.7	GrapeTree	
		In silico PCR	
		Interactive Tree of Life (iTOL)	
		Locus explorer       3         Microreact       3	
		PhyloViz	
		Polymorphisms	
		Species identification	
		Sequence bin breakdown	
		Two field breakdown	
		Unique combinations	
15	Data	export plucing	85
15		export plugins         3           Contig export         3	
		Isolate record export	
		Profile export	
		Sequence export	
16			97 07
		Registering a user account    34      Allele submission    33	
		Profile submission	
		Isolate submission	
		Genome submission	
	16.6		14
			17
17	RES	Iful Application Programming Interface (API)     4	19
17	17.1		19
	17.2		19
	17.3		$\frac{1}{20}$
	17.4		51
18	Free	aently asked questions (FAQs) 4	57
10	-		57 57
			57 58
			58
19	<b>Appe</b>		<b>61</b> 61
	19.1		01

					4 	
20	Datab	base schema			4	65
Ind	lex				4	67

#### 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

### ONE

### **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

TWO

### **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::Map
- 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.
- PDL Required for LINcode calculations.
- Time::Duration Used by Job Viewer to display elapsed time in rounded units.
- TOML Used to define dashboard layouts.
- 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.
- PostGIS needed for geographical mapping.

CHAPTER

THREE

# 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.
- 3. 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.
- 4. 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/.
- 6. 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
```

- 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
                    a11
                                                        ident map=mymap
# IPv4 local connections:
                                 127.0.0.1/32
                                                        md5
host
        all
                    a11
# IPv6 local connections:
host
                    a11
        a11
                                 ::1/128
                                                       md5
```

# MAPNAME	SYSTEM-USERNAME	PG-USERNAME
mymap	postgres	postgres
mymap	webmaster	postgres
mymap	www-data	apache
mymap	bigsdb	bigsdb
mymap	bigsdb	apache

### 3.2.2 pg\_ident.conf

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 () followed by a URL for that link, e.g.

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

Global announcements can be made in a banner that appears on each database contents page. This is useful for service announcements such as for planned maintenance. Place a HTML file called announcement.html in /etc/bigsdb including the text that you wish to appear.

### 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 (|) 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 XXXXXXX 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  $1.6 \rightarrow 1.8$ , requires upgrading to 1.7 first.

Minor version changes, e.g. 1.8.0 -> 1.8.1, have no modifications to the database structures. There will be changes to the Perl library modules and possibly to the contents of the Javascript directory, images directory and CSS files.

### 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):

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
[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
```

(continues on next page)

(continued from previous page)

#### </VirtualHost>

You should also set 'rest\_behind\_proxy=1' in bigsdb.conf.

#### CHAPTER

FOUR

# 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 as well as country and year. To add new fields, you need to log in to the database and alter this table. For example, to add fields for age and sex, 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 age int; ALTER TABLE isolates ADD sex text;

If you want to use the geography\_point field type, used for storing and mapping GPS coordinates, then you will need to install the PostGIS module for PostgreSQL and enable this within the database as follows:

CREATE EXTENSION postgis;

To create a geography\_point field for location alter the isolate table as below:

ALTER TABLE isolates ADD location geography(POINT,4326);

[SRID 4326 represents spatial data using longitude and latitude coordinates on the Earth's surface.]

Fields can also be linked to a GPS lookup table and can then be mapped. If this is enabled then you need to ensure that PostGIS is installed and create the lookup table by running the isolatedb\_geocoding.sql script against the isolate database with the following:

psql -f isolatedb\_geocoding.sql bigsdb\_test\_isolates

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 optional files:

- 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 autoregistration is not enabled).
- registration.html HTML markup for text that will appear on the login page for the current database. This appears right before the "Log in" button.

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'.
- alternative\_codon\_tables
  - Enable alternative codon tables: either 'yes' or 'no'. Set to 'yes' to allow individual isolates to use a different codon table than the default (defined by the 'codon\_table' attribute), default is 'no'.
- annotation
  - Semi-colon separated list of accession numbers with descriptions (separated by a |), 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\_table
  - Set the id of the global codon table to use. See https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi
    for a list with the of ids and their description. This can be overridden on a per-isolate basis if alternative\_start\_codons is set to 'yes'. Default value is "11".
- 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'.
- country\_field
  - Sets the field in which country is stored. Default: 'country'. This is needed for mapping.
- 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 famicroscope'.
- 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 () and an icon class from the FontAwesome library, e.g. 'Vaccine reactivity|fas 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'.
- field\_groups
  - Comma-separated list of category names that standard isolate fields can be grouped in to in the isolate information page. You can add an icon to appear by following the name with a pipe symbol (|) and an icon class from the FontAwesome library, e.g. 'Antimicrobial resistance/fas fa-capsules'.
- 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'.
- genome\_submissions
  - Enable genome submissions (automated submission system): either 'yes' or 'no', default 'yes'.
  - To enable, you will also need to set submissions="yes". By default, genome submissions are enabled.
- 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. The plugin also requires microreact\_token to be provided in bigsdb.conf.
- microreact\_country\_field
  - Overrides the field in which country is stored. Default: 'country'
- 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).

- microreact\_year\_field
  - Overrides the field in which year is stored. Default: 'year'
- 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 '|' and the description, e.g. 'pubmlst\_neisseria\_seqdef|Typing'. This is used to populate the menu items.
- 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'.
- script\_path\_includes
  - Partial path of the bigsdb.pl script used to access the database. See *user authentication*.
- separate\_dataset
  - Treat database configuration as though it was a separate database for the purposes of handling submissions and curators: either 'yes' or 'no', default 'no'. Submissions will be tagged with the configuration name and will only be visible within that configuration. Curators can be limited to specific configurations by populating the curator\_configs table. This also affects whether they are notified of submissions.
- 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\_codons
  - Semi-colon separated list of start codons to allow. Note that this list will replace the built-in defaults of ATG, GTG, and TTG, and is used for all functions that require recognising complete coding sequences, such as automated allele definition.
- 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.
- warn\_max\_contigs
  - Set a threshold for the number of contigs in a submitted genome assembly to trigger a warning in the submission interface. This value overrides the value set in bigsdb.conf.
- warn\_max\_total\_length
  - Set an upper threshold for the total size of a submitted genome assembly to trigger a warning in the submission interface.
- warn\_min\_n50
  - Set a threshold for the minimum N50 value in a submitted genome assembly to trigger a warning in the submission interface. This value overrides the value set in bigsdb.conf.
- warn\_min\_total\_length
  - Set a lower threshold for the total size of a submitted genome assembly to trigger a warning in the submission interface.

- 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>
</optlist>
</field>
```

• type

- Data type: int, text, float, bool, date, or geography\_point.
- 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.
- geography\_point\_lookup
  - Set to 'yes' if this field should be linked to a lookup table of GPS coordinates in order to facilitate mapping
    within the isolate information page and the Field Breakdown plugin. If any fields have this value set, you

need to install the lookup tables by running the isolatedb\_geocoding.sql script against the isolate database. This also requires that the PostgreSQL PostGIS module is installed.

- group
  - Fields can be grouped in the isolate information page by specifying the group attribute. The group name
    must be defined in the field\_groups system attribute, otherwise the field will not be shown at all. If undefined, the field will be in the default provenance/primary metadata group.
- hide
  - Completely ignore field. This is useful if you access a database using different configuration files and a field is not relevant to a particular instance. See also *Over-riding values*.
- isolate\_display
  - Set to 'no' to not show field on the isolate information page.
- 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.
- prefixes
  - Sets the name of a field that this field should be used as a prefix for. That field must be defined. An example of where this would be useful is for defining AMR fields, where one field is a modifier (>,<,=) for a MIC value field. A field with this attribute defined will not be shown as a separate field within the isolate record, but will be displayed as a prefix to the value of the set field. The prefix field will also not be labelled in the</p>

curation interface isolate add/update form, but will appear immediately before and inline with the prefixed field.

- query
  - Set to 'no' to exclude field from query drop-down lists.
- 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.
- separator
  - Optional string to place between field prefix value and field value if the prefixes attribute is defined.
- suffix
  - Optional string that is displayed after value in isolate information page and curation interface. Useful for adding units for numerical values.
- 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'.
- alternative\_codon\_tables
  - Enable alternative codon tables: either 'yes' or 'no', default is 'no'. Set to 'yes' to allow different codon tables to be selected when viewing translated sequences or filtering by CDS when uploading new allele definitions.
- 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'.
- codon\_table
  - Set the id of the global codon table to use. See https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi
    for a list with the of ids and their description.
- 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**\*\*.
- genome\_submissions
  - Enable link to genome submissions (automated submission system): either 'yes' or 'no', default 'yes'.
  - To enable, you will also need to set isolate\_submissions="yes".
- 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 menu items.
- script\_path\_includes

- Partial path of the bigsdb.pl script used to access the database. See user authentication.
- separate\_dataset
  - Treat database configuration as though it was a separate database for the purposes of handling submissions and curators: either 'yes' or 'no', default 'no'. Submissions will be tagged with the configuration name and will only be visible within that configuration. Curators can be limited to specific configurations by populating the curator\_configs table. This also affects whether they are notified of submissions.
- 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\_codons
  - Semi-colon separated list of start codons to allow. Note that this list will replace the built-in defaults of ATG, GTG, and TTG, and is used for all functions that require recognising complete coding sequences.
- 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 allow\_submissions, required, maindisplay, default, hide, 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.

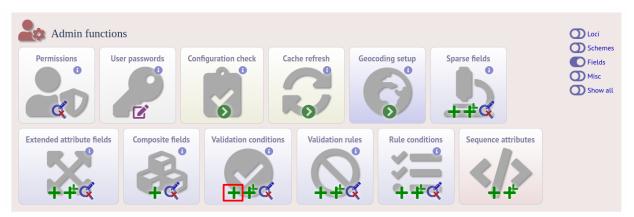
```
date_received:required="yes"
```

# 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:

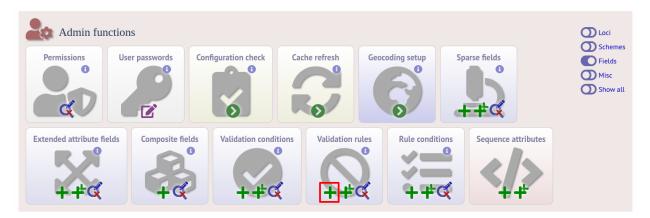


Add the following conditions separately:

- $age_year > 0$
- age\_month NOT null

PubMLST		
Home > Organisms > Organism > Neisseria isolates > Add new validation condition		
Add new validation condition		8
Please fill in the fields below - required fields are marked with an exclamation mark (!).	Action RESET SUBMIT	ł

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



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	
Please fill in the fields below - required fields are marked with an exclamation mark (!).  Record  id: 1  message: 1  ge_mth should only be set when age_yr < 1.]  curator:! Keith Jolley (keith)  datestamp: 1 2020-07-18	

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

Admin functio	ons					O Loci
Permissions U	Jser passwords	onfiguration check	Cache refresh	Geocoding setup	Sparse fields	Field Misc Show
Extended attribute fields	Composite fields	Validation conditio	NS Validation ru			

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 1) age_mth should only be set when age_yr < 1.  condition id:1 1) age_yr > 0  curator:1 Keith Jolley (keith) datestamp:1 2020-07-18	Action <b>RESET</b> SUBMIT	ľ

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]

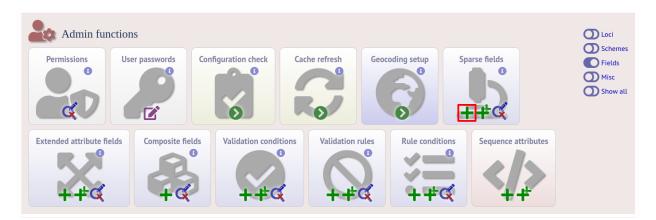
The two fields have to be of the same data type in order to be compared (you cannot compare a text field to an integer field for example).

# 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'.

PubMLST	
Home > Organisms > Organism > Neisseria isolates > Add new sparse field	
	2

#### Add new sparse field

Record			<i></i> /	Action	
	cigarettes_smoked_per_day	Â		RESET SUBMIT	
value format:!	text ✓ ○ true ● false (j)				-
	Otrue I false				
	Keith Jolley (keith)				
	Risk factors				
description:					
length:	()				
-	none;1-5;6-10;11-20;21+				
			(j)		
value regex:		()	•		
conditional formatting:					
			(i)		
html link text:	<u>(</u> )				
html message:					
min value:	<b>i</b>		J		
max value:					
field order:					

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 (|) 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
	å	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).:

INSERT INTO users (id, user\_name, surname, first\_name, email, affiliation, status, date\_ →entered, datestamp, curator) VALUES (1, 'keith', 'Jolley', 'Keith', 'keith.jolley@zoo.ox.ac.uk', 'University of Oxford, UK', 'admin', 'now', 'now', 1);

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

## 4.11 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).

PubMLST		
Home > Organisms > Organism > Leptospira isolates > Add new OAuth credentials		
Add new OAuth credentials		
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record	Action	<b>14</b>
<pre>base uri! <u>https</u>://rest.pubmlst.org/db/pubmlst_rmlst_isolates</pre>	RESET	
consumer key: RylqmtjBbSAkyyrZ5ZUyHzae		
consumer secret: U6B_^%HJP9N^1Ax2wemPPkmqrt_J%m_rX6pkgkfdho		
access token: Hent Hent Hent Hent Hent Hent Hent Hent		
access secret:   QODGPqKoSgVcGwdlKBejufDsn4B1beX1		
curator:! Keith Jolley (keith)		
date entered:! 2020-07-18		
datestamp:! 2020-07-18		

### 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).
```

```
(continues on next page)
```

```
--min ID
Minimum isolate id.
--max ID
Maximum isolate id.
--projects LIST
Comma-separated list of project isolates to scan.
--quiet
Only display errors.
.._setup_dashboard:
```

## 4.13 Setting up front-end and query dashboards

Dashboards can be used as an alternative front-end to isolate databases. They can also be used to summarise the results of a query. In order to enable dashboards for a particular database, they have to be enabled either globally or specifically for the database configuration. If enabled, users will have the option to toggle between the dashboard and the standard index page.

To enable globally and use the front-end dashboard by default, set the following in bigsdb.conf:

enable\_dashboard=1
default\_dashboard\_view=1

Each of these values can be overridden for a particular database by setting the same attribute in the database config.xml file, with either 'yes' or 'no', i.e. dashboards can be enabled globally but disabled for a particular database configuration, or disabled globally but enabled for a particular database configuration.

To enable query dashboards, set the following in bigsdb.conf:

```
query_dashboard=1
```

Again, this value can be overridden for a particular database by setting the attribute in the database config.xml file, as above. Note that 'enable\_dashboard' also needs to be enabled.

#### 4.13.1 Defining default dashboards

A default global front-end dashboard can be set up by placing a dashboard\_primary.toml file in /etc/bigsdb. This can be overridden for individual database configurations by adding a TOML file (dashboard\_primary.toml), in the same format, to the database configuration directory. Any field defined in the TOML file that does not appear within a particular database is ignored.

A default global query dashboard can be similarly set up with a file called dashboard\_query.toml. As this dashboard appears above isolate results tables it is preferred that smaller elements are defined, usually with a height of 1.

An example of the format can be seen below.

```
#Configuration for default front-end dashboard for isolate databases. This
#defines the visual elements that will be included. If field-specific elements
```

```
#are defined and that field does not exist in a particular database then it
#will be ignored.
#The default configuration can be overridden for a particular database by
#including a dashboard.toml file, using the same format, in the database
#configuration directory.
#Width can be 1, 2, 3, or 4.
#Height can be 1, 2, or 3.
#Field names have prefixes indicating the field type:
#f_ are standard provenance/primary fields
#e_ are extended attributes with the main field and the attribute separated
  by //, e.g. e_country//continent.
#
elements = \Gamma
  { #Isolate count.
     display = 'record_count',
                     = 'Isolate count',
     name
     width
                     = 2,
     background_colour = '#79cafb',
     main_text_colour = '#404040',
     watermark
                  = 'fas fa-bacteria',
     change_duration = 'month',
                     = 'Browse isolates',
     url_text
     hide_mobile
                     = 0
  }.
  { #Genome count (will only display if there are genomes in the database).
     display = 'record_count',
                      = 'Genome count',
     name
     genomes
                     = 1,
     width
                     = 2,
     background_colour = '#7ecc66',
     main_text_colour = '#404040',
     watermark = 'fas fa-dna',
     change_duration = 'month',
                  = 'Browse genomes',
     url_text
     post_data
                    = \{ \text{genomes} = 1 \},
     hide_mobile
                     = 0
  },
  {
                      = 'field',
     display
                      = 'Country',
     name
                      = 'f_country',
     field
     breakdown_display = 'map',
     width = 3,
     height
                      = 2,
     hide_mobile
                      = 1
  },
  {
     #Top 5 list of continents (Geocoding should be set up with default country
      #list linked to continent - see 'Geocoding setup' on admin curator page.
```

```
display
              = 'field',
                 = 'Continent',
  name
  field
                  = 'e_country||continent',
  breakdown_display = 'top',
  top_values
                = 5,
  width
                  = 2.
  hide_mobile
                  = 1
},
{
                  = 'Sequence size',
  name
                  = 'seqbin_size',
  display
  genomes
                = 1,
  hide_mobile
                = 1,
  width
                  = 2,
  height
                  = 1
},
   #Doughnut chart of species.
{
  display = 'field',
  name
                 = 'Species',
         - open
= 'f_species',
  field
  breakdown_display = 'doughnut',
  height = 2,
  width
                 = 2,
  hide_mobile
                 = 1
},
   #Treemap of disease.
{
  display = 'field',
  name = 'Disease',
field = 'f_disease',
                 = 'Disease',
  breakdown_display = 'treemap',
  height = 2,
                 = 2,
  width
  hide mobile
                  = 1
},
{
  #Bar chart of submission years.
  display = 'field',
  name
field
                 = 'Year',
                 = 'f_year'.
  breakdown_display = 'bar',
  width = 3,
  bar_colour_type = 'continuous',
  chart_colour = '#126716',
hide_mobile = 1
  hide_mobile
},
{
  #Cumulative chart of submissions by date.
  display = 'field',
  name
                 = 'Date entered',
  field
                 = 'f_date_entered',
  width
                  = 2,
  breakdown_display = 'cumulative',
```

```
hide_mobile = 1
}
```

#### **Attributes**

The allowed attributes are listed below.

- background\_colour
  - RGB hex code for the background colour, e.g. '#79cafb'. This is used only for 'big number' fields, e.g. isolate count.
- bar\_colour\_type
  - categorical use contrasting colours for bars.
  - continuous use the same colour for all bars (set colour use 'chart\_colour' attibute).
- breakdown\_display type of visualisation. Allowed values are:

– bar

- \* bar chart particularly useful for continuous data such as year.
- cumulative
  - \* cumulative line chart used for date\_entered or datestamp fields.
- doughnut
  - \* doughut chart
- gps\_map
  - \* GPS map. This can only be used for geography\_point fields or fields that are linked to a lookup table of GPS coordinates.
- pie

```
* pie chart
```

- top
  - \* top values list. You can choose the number of values to display by also setting the top\_values attributes to either 3, 5, or 10.
- treemap
  - \* treemap chart

- word

\* word cloud. This can only be used for fields that have a defined list of allowed values.

– map

- \* global map. This can only be used for 'country' fields with a defined list allowed values or 'continent' fields which are an extended attribute of country.
- change\_duration
  - Show the rate of change, e.g. the number of new records in past month. Used for 'big number' fields, e.g. isolate count or specific value count. Allowed values are 'week', 'month', or 'year'.

- chart\_colour
  - RGB hex code for bar or cumulative line charts, e.g. '#126716'.
- display
  - Element type. Allowed values are:
    - \* field This is used for most elements.
    - \* record\_count Used for isolate count fields.
    - \* seqbin\_size Used to display a histogram of genome sizes.
- field
  - The name of the field to display. Different types of field have different prefixes as follows:
    - \* Primary isolate field prefix with 'f\_', e.g. 'f\_country'.
    - \* Secondary metadata fields prefix with 'eav\_'.
    - \* *Extended attributes* prefix with 'e\_', followed by the primary field name, followed by '||' and then the extended attribute name, e.g. for continent linked to country you would use 'e\_country||continent'.
    - \* Scheme fields e.g. clonal complex prefix with 's\_' followed by the scheme id number, then '\_', followed by the scheme field name, e.g. for a field called 'clonal\_complex' defined for scheme 1, you would use 's\_1\_clonal\_complex'.
- gauge\_background\_colour
  - RGB hex code, e.g. '#79cafb, for the background colour on a gauge chart.
- gauge\_foreground\_colour
  - RGB hex code, e.g. '#79cafb, for the foreground colour on a gauge chart.
- geography\_view
  - Choice of view for GPS maps, either 'Aerial' or 'Map'. Note that Aerial views can only be used if you have a Bing Maps key set in bigsdb.conf.
- header\_background\_colour
  - RGB hex code, e.g. '#79cafb#, for the header background for a top values list.
- header\_text\_colour
  - RGB hex code, e.g. '#79cafb#, for the header text colour for a top values list.
- height
  - Height of element either 1, 2, or 3.
- hide\_mobile
  - Set to 1 to hide element on small mobile devices (width <= 480 pixels).
- main\_text\_colour
  - RGB hex code, e.g. '#79cafb#, for the colour of the text used in big number elements.
- marker\_colour
  - RGB hex code, e.g. '#79cafb#, or HTML colour value, for the colour of the markers on GPS maps.
- marker\_size
  - Size of marker on GPS maps. Allowed values are 0-9.

- max\_contigs
  - Number of contigs above which a genome will be rejected in the submission interface.
- min\_n50
  - N50 value below which a genome will be rejected in the submission interface.
- min\_total\_length
  - Minimum length in base pairs below which a genome will be rejected in the submission interface.
- name
  - The name used for the title of the element.
- palette
  - ColorBrewer palette used for map displays. Allowed values are:
    - \* blue
    - \* green
    - \* purple
    - \* orange
    - \* red
    - \* blue/green
    - \* blue/purple
    - \* green/blue
    - \* orange/red
    - \* purple/blue
    - \* purple/blue/green
    - \* purple/red
    - \* red/purple
    - \* yellow/green
    - \* yellow/green/blue
    - \* yellow/orange/brown
    - \* yellow/orange/red
- post\_data
  - Used to pass data attributes for linked queries. Currently only 'genomes' is used to specify that isolates should be filtered to those with genome assemblies, e.g. '{genomes = 1}'.
- specific\_value\_display
  - Type of display to use for specific values. Allowed values are:
    - \* gauge gauge chart
    - \* number big number value
- specific\_values
  - list of field values to include in count shown in gauge chart or big number display, e.g. '['Neisseria meningitidis']'.

• url\_text

- link text to display when hovering over link leading to data query. Only available for isolate count or specific value charts.
- visualisation\_type
  - Either 'breakdown' (default) or 'specific values'. You need to then set the visualisation using either the breakdown\_display or specific\_value\_display attribute.
- warn\_max\_contigs
  - Number of contigs above which a genome will have a warning message shown in the submission interface.
- warn\_min\_n50
  - N50 value below which a genome will have a warning message shown in the submission interface.
- warn\_min\_total\_length
  - Minimum length in base pairs below which a genome will have a warning shown in the submission interface.
- watermark
  - FontAwesome icon class used for background watermark on big number charts, e.g. 'fas fa-bacteria'. See https://fontawesome.com/icons?m=free.
- width
  - Width of element either 1, 2, 3, or 4.

#### CHAPTER

# **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\_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).
- modify\_geopoints allowed to modify the geography point GPS coordinate lookup table used for mapping.

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	elp 🔽
Select curator(s)     Action       Debech, Nadia (ndebech)     Deghmane, Ala-Eddin (deghmane)       Demczuk, Walter (wdemczuk)     Demczuk, Walter (wdemczuk)       Desmet, Stefanie (sdsmet6)     Diallo, Kanny (kdiallo)       Dildion, Jo-Anne (dillon)     Dillon, Jo-Anne (dillon)       Exley, Rachel (rexley)     Image: Comparison of the second s	

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

MLST									
e > Organisms > Orga	nism > Neisseria isol	ates > Set cu	rator permis	ions					
									Help
t curator pe	ermissions								
1									
Select curator(s)		Acti	on						
Cehovin, Ana (acehovi	n)	•							
Chan, Hannah (hchan)		SEL	LECT						
Chen, Mingliang (cmlb									
Choi, Eunhwa (euncho Christodoulides, Myron									
Clark, Stephen (SClark									
Claus, Heike (hclaus)	7								
Clemence, Marianne (r	nclemence)	-							
All	None								
	autrad parmissians 11	core with a sta	tur of 'cubm				In stand Adda through a set	with a rod background add rostrig	
leck the boxes for the re	quirea permissions. O	sers with a sta	atus or subin	tter' have a restricted lis	st of allowed permiss	sions that can be se	elected. Attributes v	with a reu background add restric	ctions.
		5015 WILLI & 500	itus or subiri		st of allowed permiss	sions that can be se	lected. Attributes v	with a <mark>red background</mark> add restric	ctions.
Update permissions		sers with a sta	itus or subin	tter' have a restricted lis Action	st of allowed permiss	sions that can be se	lected. Attributes (	with a red background add restric	ctions.
Update permissions				Action	st of allowed permiss	sions that can be se	lected. Attributes \		ctions.
Update permissions	Curato Clemence, Marianne	n Diallo, Kanny	All/None		st of allowed permiss	sions that can be se	lected. Attributes v	with a red background aud restric	ctions.
Update permissions Permission query users	Curator Clemence, Marianne	r Diallo, Kanny	All/None	Action	st of allowed permiss	sions that can be se	lected. Attributes v		ctions.
Update permissions Permission query users modify users	Curator Clemence, Marianne	Diallo, Kanny	All/None	Action	st of allowed permiss	sions that can be se	lected. Attributes v	with a rea background aud result	ctions.
Update permissions Permission query users modify users modify isolates	Curato Clemence, Marianne	Diallo, Kanny	All/None	Action	st of allowed permiss	sions that can be se	lected. Attributes v	with a red background add result	ctions.
Update permissions Permission query users modify users modify isolates modify projects	Curator Clemence, Marianne	r Diallo, Kanny O C Z	All/None	Action	st of allowed permiss	sions that can be se	lected. Attributes (	with a red background add result	ctions.
Update permissions Permission query users modify users modify isolates modify projects modify sequences	Curato Clemence, Marianne	r Diallo, Kanny C C C C	All/None	Action	st of allowed permiss	ions that can be se	lected. Attributes (		ctions.
Update permissions Permission query users modify users modify folates modify projects modify sequences tag sequences	Curato Clemence, Marianne	r Diallo, Kanny C C C C C C	All/None	Action	st of allowed permiss	ions that can be se	lected. Attributes v		ctions.
Permission query users modify users modify isolates modify solates modify solates addify sequences tag sequences designate alleles	Curato Clemence, Marianne Clemence, Marianne C C C C C C	Diallo, Kanny	All/None	Action	st of allowed permiss	ions that can be se	lected. Attributes (		ctions.
Permissions Permission query users modify users modify users modify projects modify sequences tag sequences designate alleles modify usergroups	Curato Clemence, Marianne	Diallo, Kanny	All/None	Action	st of allowed permiss	ions that can be se	lected. Attributes (		ctions.
Permission Query users modify users modify isolates modify projects modify projects a gequences tag sequences designate alleles modify usergroups set user passwords	Curatoi Clemence, Marianne Clemence, Marianne C C C C C C C C C C C C C C C C C C	Diallo, Kanny	All/None	Action	st of allowed permiss	ions that can be se	tected. Attributes i		ctions.
Update permissions Permission query users modify users modify isolates modify isolates modify projects modify projects designate alleles modify usergroups set user passwords modify loci	Curato Clemence, Marianne	Diallo, Kanny	All/None	Action	st of allowed permiss	ions that can be se	tected. Attributes i		ctions.
Permissions Permission query users modify users modify users modify rojects modify rojects tag sequences tag sequences designate alteles modify usergroups set user passwords modify loci modify schemes	Curato Clemence, Marianne	Diallo, Kanny	All/None	Action	st of allowed permiss	ions that can be se	tected. Attributes i		ctions.
Update permissions Permission query users modify users modify projects modify projects modify projects agequences designate alteles modify usergroups set user passwords modify loci modify composites	Curator Clemence, Marianne Clemence, Marianne C C C C C C C C C C C C C C C C C C	Diallo, Kanny Diallo, Kanny C C C C C C C C C C C C C C C C C C	All/None	Action	st of allowed permiss	ions that can be se	lected. Attributes Y		ctions.
Update permissions Permission query users modify users modify rojects modify sequences designate alleles modify usergroups set user passwords modify loci modify schemes modify composites	Curato Clemence, Marianne	Diallo, Kanny Diallo, Kanny C C C C C C C C C C C C C C C C C C	AU/None	Action	st of allowed permiss	ions that can be se	lected. Attributes i		ctions.
Update permissions Permission query users modify users modify projects modify sequences tag sequences designate alleles modify usergroups set user passwords modify loci modify schemes modify composites modify que attributes	Curato: Clemence, Marianne	Diallo, Kanny	ALVNone	Action	st of allowed permiss	ions that can be se	tected. Attributes 1		ctions.
Update permissions Permission query users modify lsolates modify projects modify projects modify projects designate alteles modify usergroups set user passwords modify composites modify composites modify feld attributes modify value attributes	Curato Clemence, Marianne	Callo, Kanny Callo	ALL/None	Action	st of allowed permiss	ions that can be se	tected. Attributes Y		ctions.
Update permissions Permission query users modify users modify projects modify sequences designate alleles modify usergroups set user passwords modify fold modify fold attributes modify romposites modify romposites modify probes	Curato Clemence, Marianne	Diatio, Kanny	ALL/None	Action	st of allowed permiss	ions that can be se	tected. Attributes i		cuons.
Update permissions Permission query users modify users modify projects modify projects tag sequences tag sequences designate alleles modify valvergroups set user passwords modify vol- modify schemes modify composites modify composites modify field attributes modify sparse fields modify probes modify probes	Curato: Clemence, Marianne	Contraction of the second seco	All/None	Action	st of allowed permiss	ions that can be se	tected. Attributes i		ctions.
Update permissions Permission query users modify users modify isolates modify rojects modify sequences tag sequences tag sequences tag sequences tag sequences tag sequences tag sequences tag sequences modify usergroups set user passwords modify composites modify fold attributes modify fold attributes modify probes modify probes modify eperiments delete all	Curato Clemence, Marianne	Diatlo, Kanny C C C C C C C C C C C C C C C C C C	All/None	Action	st of allowed permiss	ions that can be se	tected. Attributes 1		cuons.
Update permissions Permission query users modify users modify users modify sequences designate alleles modify usergroups set user passwords modify loci modify schemes modify composites modify value attributes modify value attributes modify yarse fields modify probes modify probes modify experiments detete all import site users	Curato Clemence, Marianne	Diallo, Kanny	All/None	Action	st of allowed permiss	sions that can be se	tected. Attributes 1		cuons.
Update permissions Permission query users modify users modify solates modify projects modify projects tag sequences tag sequences tag sequences designate alleles modify valvergroups set user passwords modify valvergroups set user passwords modify composites modify field attributes modify field attributes modify probes modify probes modify probes modify probes modify spreeriments delete all import site users	Curato Clemence, Marianne	Diallo, Kanny	All/None	Action	st of allowed permiss	ions that can be se	tected. Attributes 1		ctions.
Update permissions Permission query users modify users modify users modify sequences designate alleles modify usergroups set user passwords modify loci modify schemes modify composites modify value attributes modify value attributes modify yarse fields modify probes modify probes modify experiments detete all import site users	Curato: Clemence, Marianne	Diallo, Kanny	All/None	Action	st of allowed permiss	sions that can be se	tected. Attributes 1		cuons.

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

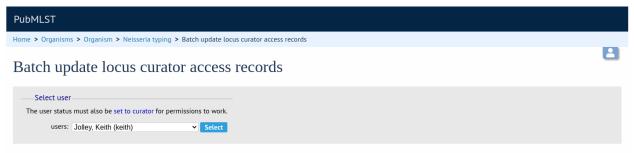
# 5.4 Locus and scheme permissions (sequence definition database)

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':

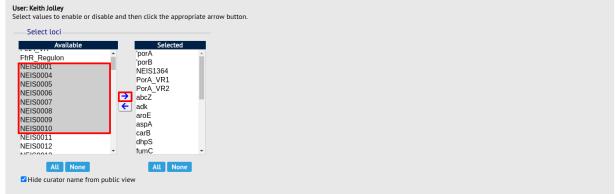


#### Select the curator from the list:



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





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

# 5.5 Controlling access

### 5.5.1 Restricting particular configurations to specific user accounts

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

In the config.xml, add the following directive:

```
default_access="deny"
```

This tells BIGSdb to deny access to anybody unless their account name appears within a file called users.allow within the config directory. The users.allow file should contain one username per line. 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	ism > Neisseria typing > Change password		
Set user passv	vord	E	3
Passwords must be at least	8 characters long.		
Passwords		Action	
User:	Jolley, Keith (keith) ~	SET PASSWORD	
New password:	••••••	SET PASSWORD	
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 = (
               name => 'BURST',
author => 'Keith Jolley',
               affiliation => 'University of Oxford, UK',
               email => 'keith.jolley@zoo.ox.ac.uk',
               description => 'Perform BURST cluster analysis on query results query_
\rightarrow results'.
               category => 'Cluster',
               buttontext => 'BURST',
               menutext => 'BURST',
               module => 'BURST',
               version => '1.0.0',
dbtype => 'isolate
                           => 'isolates, sequences'.
               section => 'postquery',
               order
                          => 10,
               system_flag => 'BURST',
                       \Rightarrow 'query',
               input
               requires => 'mogrify',
                           => 2,
               min
                            => 1000
               max
       );
       return \%att;
}
```

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

BURST="yes"

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

<pre>#Existing_host</pre>	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
```

(continued from previous page)

```
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:

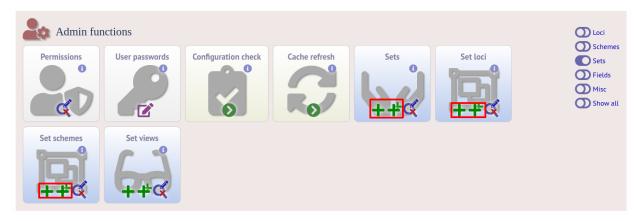
Sets (concept)

## 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 pro <b>158</b> :1005-15.	ein genes. rMLST is described in Jo	olley et al. 2012 Microbiolog	y
For academic non commercial use only. For full terms and conditions please see rMLST_licence.pdf. To discuss any other sort of us innovation@isis.ox.ac.uk quoting reference 7895/MC.	e, including commercial use licence	e, please contact	
Datasets			
This database contains multiple datasets.			
Please select: Bacterial domain			
Query database		Т	+
Search database		π	+
Browse, search by any criteria, or enter list of attributes.	🗠 ANALY	SIS	
	🕸 СИЗТО	MISE	+
Search by combinations of loci This can include partial matches to find related isolates.		MATION	+
		G	

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:

```
CREATE VIEW spneumoniae AS SELECT * FROM isolates WHERE species = 'Streptococcus∟

→pneumoniae';

GRANT SELECT ON spneumoniae TO apache;
```

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

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

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

#### 5.12.4 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.

Admin fur	nctions				Doci Schemes
Permissions	User passwords	Configuration check	Cache refresh	User databases	<ul> <li>Schemes</li> <li>Fields</li> <li>Misc</li> <li>Show all</li> </ul>

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 > Neisseria isolates > Add new user database		
Add new user database		2
Please fill in the fields below - required fields are marked with an exclamation mark (!)Record	Action	- 14
id:! 1 name:! PubMLST  Bite/domain name	RESET SUBMIT	
dbase name: [ pubmist_bigsdb_users] Name of the database holding user data		,
curator:! Keith Jolley (keith) datestamp:! 2020-07-19		
list order: auto registration: Otrue Ofalse Allow user to register themself for database		
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

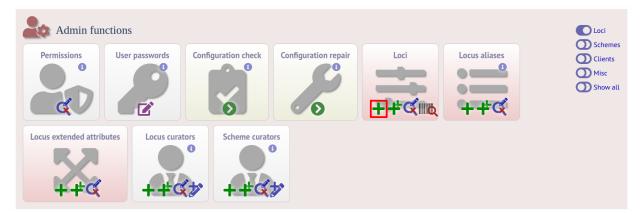
#### See also:

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:

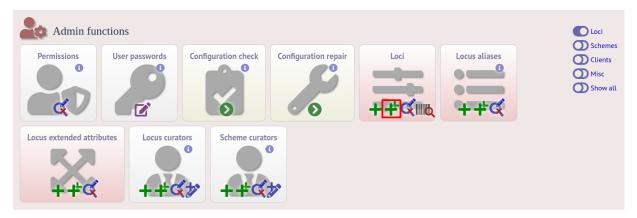
- id The name of the locus.
  - Allowed: any value starting with a letter, number or underscore.
- data\_type Describes whether the locus is defined by nucleotide or peptide sequence.
  - Allowed: DNA/peptide.
- allele\_id\_format The format for allele identifiers.
  - Allowed: integer/text.
- length\_varies 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).
- complete\_cds Whether locus represents a complete coding sequence (optional)
- start\_codons Semi-colon separated list of alternative start codons to allow
  - Note that these are in addition to the built-in defaults of ATG, GTG, TTG.
- 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.



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		2
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. <b>Templates</b> Display         Check the description of database fields for help with filling in the template. <b>Upload</b> Paste in tab-delimited text (include a field header line).	Action SUBN	

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 fur	nctions				
Permissions	User passwords	Configuration check	Cache refresh	Locus aliases	<ul> <li>Schemes</li> <li>Fields</li> <li>Misc</li> <li>Show all</li> </ul>
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 > 1	Veisseria isolates > Add new locus	
Add new locus		Show tools
Please fill in the fields below - requ	ired fields are marked with an exclamation mark (!).	
Record		
id	d 🔤	
allele id format		
	:! Otrue @ false	
coding sequence	:! @true Ofalse	
	: allele only -	
	:! Otrue • false	
	: • true O false	
	:! ● true ○ false :! Keith Jolley (keith)	
	: 2020-07-19	
	1 2020-07-19	
formatted name		
common name	a	
formatted common name		
locus type		
allele id rege		
length		
complete cd: or		
genome positior		
match longes		
reference sequence		
pcr filte	r: Otrue Ofalse Do NOT set to true unless you define PCR reactions linked to this locus.	
	r: Ortue O false Do NOT set to true unless you define probe sequences linked to this locus.	
	s: O true O false Set to true if locus can contain introns.	
	e: pubmlst_bigsdb_neisseria_seqdef Name of the database holding allele sequences t: IP address of database host	
dbase hos dbase por		
dbase use		
dbase password		
dbase io	1: PUT_LOCUS_NAME_HERE 🖏 Name of locus in seqdef database	
description ur	ا: //cgi-bin/bigsdb/bigsdb.pl?	
	db=pubmlst_neisseria_seqdef&page=locusInfo&locus=PUT_LOCUS_NAME_HERE	
ur	<sup>l:</sup> /cgi-bin/bigsdb/bigsdb.pl? db=pubmlst neisseria seqdef&page=alleleInfo&locus=PUT LOCUS NAME HERE&allele	
	id=[?]	
submission template		
view		
aliases		
Action		
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.
- complete\_cds Whether locus represents a complete coding sequence (optional)
- · start\_codons Semi-colon separated list of alternative start codons to allow

- Note that these are in addition to the built-in defaults of ATG, GTG, TTG.
- 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.

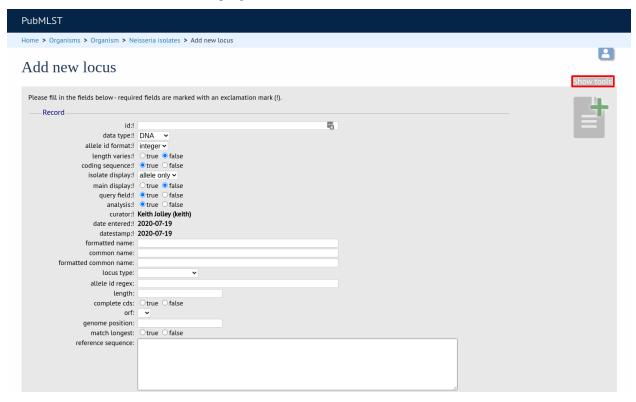
• url - The URL used to hyperlink to information about the allele. This can either be a relative or absolute URL. If [?] (including the square brackets) is included then this will be substituted for the allele value in the resultant URL. To link to the appropriate allele info page on a corresponding 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:



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

PubMLST		
Home > Organisms > Organism > Neisseria isolates > Add new	locus	
		8
Add new locus		
		Hide tools
Please fill in the fields below - required fields are marked with ar	exclamation mark (!).	Copy configuration from NEIS0001 (lpxC)
		All parameters will be copied except id, common name, reference sequence, genome
Record		position and length. The copied locus id will be substituted for 'PUT_LOCUS_NAME_HERE'
id:!		in fields that include it.
data type:! DNA 🗸		
allele id format:! integer 🗸		
length varies:! Otrue 🖲 false		
coding sequence:   true  false		
isolate display:! allele only 🗸		
main display:! 🔿 true 🖲 false		
query field:   true  false		
analysis: 1 🔍 true 🔿 false		
curator:! Keith Jolley (keith)		
date entered:! 2020-07-19		
datestamp:! 2020-07-19		
formatted name:		
common name:		
formatted common name:		
locus type: 🗸 🗸		
allele id regex:		
length:		
complete cds: Otrue Ofalse orf: V		
genome position:		
match longest: Otrue Ofalse reference sequence:		
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:	$\bigcirc$ true $\bigcirc$ false	Do NOT set to true unless you define PCR reaction	s linked to this	s locus.		
probe filter:	⊖true ⊖false	○ true ○ false Do NOT set to true unless you define probe sequences linked to this locus.				
introns:	$\bigcirc$ true $\bigcirc$ false	Set to true if locus can contain introns.				
dbase name:	pubmlst_bigsdb_nei	isseria_seqdef		Name of the database holding allele sequences		
dbase host:			IP address of	f database host		
dbase port:		Network port accepting database connecti	ons			
dbase user:		5				
dbase password:						
dbase id:	PUT_LOCUS_NAM	E_HERE	Name of locu	us in seqdef database		
description url:	/cgi-bin/bigsdb db=pubmlst_neis	o/bigsdb.pl? sseria_seqdef&page=locusInfo&locu:	=PUT_LOCU	US_NAME_HERE		
url:	/cgi-bin/bigsdb db=pubmlst_neis _id=[?]	o/bigsdb.pl? sseria_seqdef&page=alleleInfo&locu	is=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	octions				Loci
Permissions	User passwords	Configuration check	Cache refresh	Locus aliases	<ul> <li>Schemes</li> <li>Fields</li> <li>Misc</li> <li>Show all</li> </ul>
PCR reactions	PCR locus links	Nucleotide probes			

Click the link to download an Excel template:

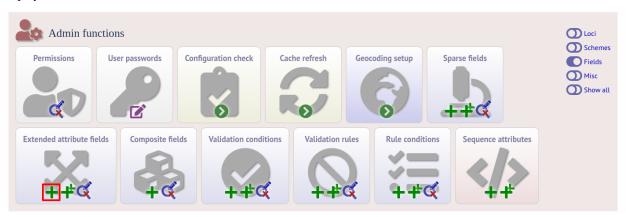
PubMLST		
Home > Organisms > Organism > Neisseria isolates > Batch add locus records		
Batch insert loci		<b>±</b>
Instructions The page allows you to upload locus data as tab-delimited text or copied from a spreadsheet.  i. elid header names must be included and fields can be in any order. Optional fields can be omitted if you wish.  i. ther aliases (alternative names) for your locus as a semi-colon (;) separated list. <b>Destination</b> Complete  Complete  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 > C	rganism > Neisseria isolates > Add new isolate field extended attribute		
Add new is	olate field extended attribute		2
Please fill in the fields Record	below - required fields are marked with an exclamation mark (!).	Action	- H-
isolate field:! attribute:!	<b>▼</b>	RESET SUBMIT	E
	text ~ 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 (|) 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.

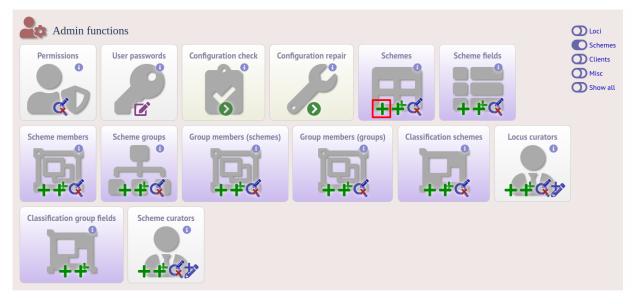
#### See also:

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.

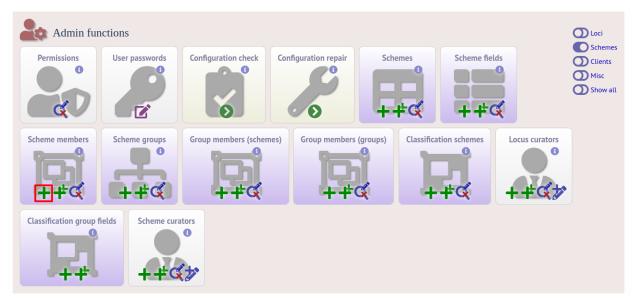


Fill in the scheme description in the web form. The next available scheme id number will be filled in already.

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 (!).  Record  id: 1  name: MLST  curator:! Keith Jolley (keith)  datestamp:! 2020-07-19  date entered: 12020-07-19	
description:	
display order:	
allow missing loci: Orue Ofalse () This is only relevant to schemes with primary key fields, e.g. MLST.	
max missing:	
disable: O true O false Set to true to disable scheme. This can be overridden by user preference settings. no submissions: O true O false Set to true to prevent submission of profiles of this scheme via the automated submission system.	
flags: experimental in development please cite unpublished - Use CTRL/SHIFT click to select or deselect values	
PubMed ids:	
links: (Format: URL/description)	

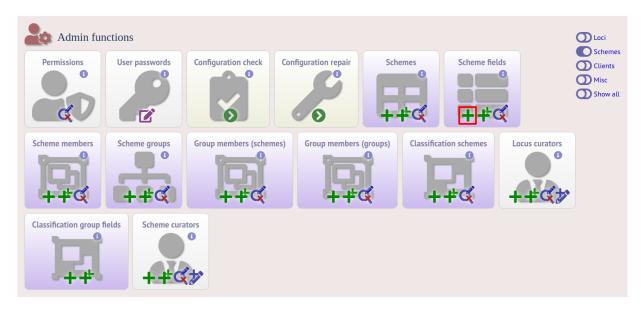
To add loci to the scheme, click the add (+) scheme members link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.



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					
Please be aware 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 aff any profiles will have to be reloaded.	ect allele designations, but				
Please fill in the fields below - required fields are marked with an exclamation mark (!).  Record Action	-				
scheme id: MLST  Coust: abcZ Curator: Keth Jolley (keth)					
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 >	> Organism > Neisseria typing > Add new scheme field	
Add new s	scheme field	3
Please be aware tha any profiles will hav	hat 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 designations, t ave to be reloaded.	but
Please fill in the fiel	ields below - required fields are marked with an exclamation mark (!).	le.
Record	Action	$\mathbf{t}$
scheme id:!	dd RESET SUBMIT	
field:		
type:!		
	y! Orus @faise ()	
	nd Otrue @false @	
	r:: Keith Jolley (keith) p:: 2020-07-19	
description:		
field order:		
index:	ex: ○true ○ false ①	
value regex:	ex. (1)	

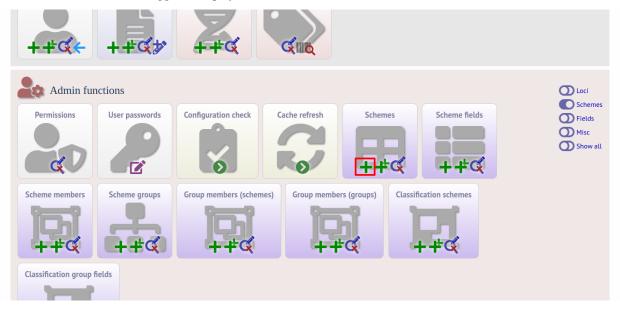
- scheme\_id Dropdown box of scheme names.
  - Allowed: selection from list.
- field Field name.
  - Allowed: any value.
- type Format for values.
  - Allowed: text/integer/date.
- primary\_key Set to true if field is the primary key. There can only be one primary key for a scheme.
  - Allowed: true/false.
- 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	n > Neisseria isolates > Add new scheme	
Add new schen		
Record id: name: isolate display: main display: query field: query status: analysis:		ŧ
dbase name:	0	
dbase host:	0	
dbase port:	0	
dbase user:	0	
dbase password:	0	
dbase id:	0	
view:	· ()	
display order:		
	Otrue of false ()       This is only relevant to schemes with primary key fields, e.g. MLST.         experimental in development       Image: Comparison of the scheme schem	
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).

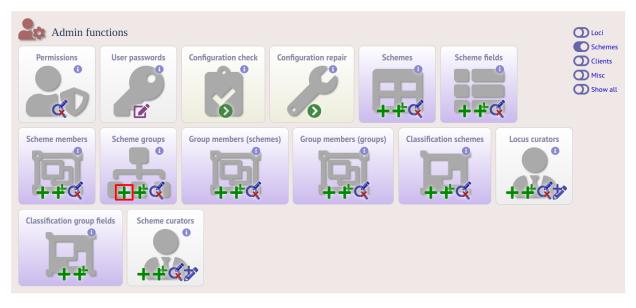
# 5.17 Organizing schemes into hierarchical groups

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 > Orgar	isms > Org	anism	> Neisseria typin	g > Downlo	oad alleles	
Downl	oad al	lele	e sequen	ces		Help Z
Select loci b	scheme   A	Inhabet	tical list   All loci l	ov scheme		
	he tree to d	isplay d		·	nemes or groups o	of schemes - clicking a group folder will display the loci for all schemes within the group and any subgroups. Click the
	enetic Inforr enomic islar neage Scher etabolism gonorrhoea asmids	ids mes ie AMR gantige	-			-
MLST						
abcZ			Length (setting) Fixed: 433 bp	Min length 433	Max length Full r 434	name/product Curator(s) Last updated O. Harrison, K. Jolley 2020-06-23
adk 🛃	DNA		Fixed: 465 bp	464	465	O. Harrison, K. Jolley 2020-07-14
aroE	DNA		Fixed: 490 bp	489	493	O. Harrison, K. Jolley 2020-06-12
fumC		1184		463	467	O. Harrison, K. Jolley 2020-06-12
gdh a			Fixed: 501 bp Fixed: 480 bp	501 480	513 503	0. Harrison, K. Jolley 2020-07-03 0. Harrison, K. Jolley 2020-06-04
pdhC pgm			Fixed: 480 bp	480	489	O. Harrison, K. Jolley 2020-06-04 O. Harrison, K. Jolley 2020-06-12
	×					

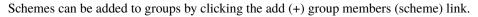
Scheme groups can be added in both the sequence definition and isolate databases. To add a new group, click the add (+) scheme group link on the curator's contents page. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it.

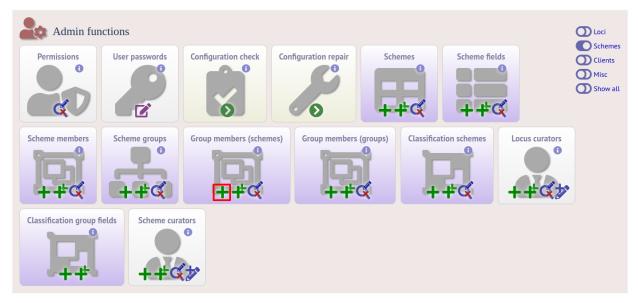


Enter a short name for the group - this will appear within drop-down list boxes and the hierarchical tree, so it needs to be fairly short.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Add new scheme group	
Add new scheme group	
Please fill in the fields below - required fields are marked with an exclamation mark (!).	
Record Action	
id: 1 RESET	
name: Typing	
curator:! Keith 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.





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

PubMLST		
Home > Organisms > Organism > Neisseria typing > Add new scheme gr	roup scheme member	
Add new scheme group scheme me	ember	20
Please fill in the fields below - required fields are marked with an exclamation of the fields are marked with are	ation mark (!). Action RESET SUBMIT	Ħ

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

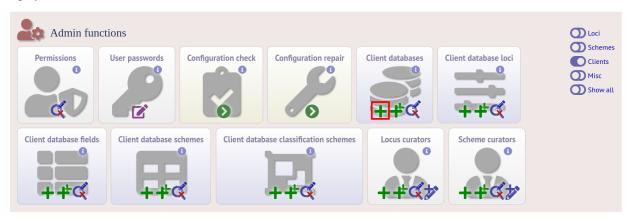
# 5.18 Setting up client databases

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:

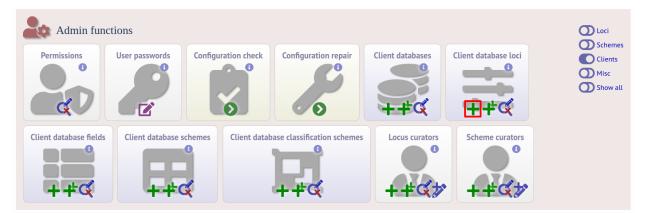
PubMLST					
Home > Organisms > Organ	ism > Neisseria typing > Add new client database				
Add new clier	nt database		<b>2</b> 0		
Please fill in the fields belo	w - required fields are marked with an exclamation mark (!).		÷		
id:!	1 PubMLST isolates				
	Contains data for a collection of isolates that represent the total known diversity of <u>Neisseria</u> species. For every <u>allelic</u> 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.				
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:!					
dbase host:		IP address of database host			
dbase port:	Network port accepting database control	inections			
dbase user:		()			
dbase password:		1			
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 > Organism > Neisseria typing > Add new locus to client database definition Add new locus to client database definition Please fil in the fields below - required fields are marked with an exclamation mark (!). Record client dbase id: 1) PubMLST isolates > locus: 1 abc2 curator: 1 Kethi Jolley (keth) datestampi: 2020-07-19 locus alias: Teset SUBMIT

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 > Organis	m > Neisseria typing > Allele information						
Allele informa	Allele information - abcZ: 5						
allele:	abcZ 5 TITGATACCG TTGCCGAAGG TTTGGGCGAA ATTCGCGATT TATTGCGCCG TTATCATCAT GTCAGCCATG AGTTGGAAAA TGGTTCGAGT GAGGCTTTGT TGAAAGAGACT TAACGAATTG CAACTTGAAA TCGAAGCGAA GGACGGCTGG AAGCTGGATG CGCGCAGTCAA GCAGACTTG GGTGAACTTG GTTGCCAGA AAACGAAAAA ATCGGCAACC TCTCCGGCGG ACAGAAAAG CGTGTGCCC TAGCGCAGGC TTGGGTGCAG AAGCCTGATG TATTGCTGCT GGACGAACCG ACCAACCATT GGACATTGA CGCGATTATT TGGCTGGAAA ATCTGCTTAA AGCGTTTGAA GGCAGCCTGG TTGTGATTAC CCACGACCGC CGTTTTTTGG						
date entered: datestamp: sender:	Sanger trace checked						
	ning this allele 302 profiles ses						
	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. 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.



## 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.

Admin funct	tions					
Permissions	User passwords	aration check	Configuration repair	Client databases	Client database loci	<ul> <li>Schemes</li> <li>Clients</li> <li>Misc</li> <li>Show all</li> </ul>
Client database fields	Client database schemes	Client datab	ase classification schemes	Locus curators	Scheme curators	

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 definiti	on				
Add new locus to client database isolate field definition					
Please fill in the fields below - required fields are marked with an exclamation mark (!).    Record  client dbase id: 1) PubMLST isolates   locus: penA  isolate field: penicillin_range  curator: Keith Jolley (keith)  datestamp: 2020-07-19  allele query: • true O false set to true to display field values when an allele query is done.	Action RESET SUBMIT	Ħ			

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 > Organism > Neisseria typing > Allele information
Allele information - penA: 9
Provenance/meta data
locus: penA
allele: 9
sequence: GACGGCGTTT TGCTGCCGGT CAGCTTTGAA AAACAGGCGG TTGCGCCGCA AGGCAAACGT ATATTTAAAG CATCGACCGC ACGTCAGGTG CGTGAGTTGA TGGTTTCTGT AACCGAACCT GGCGGTACGG GTACGGCGGT GTGCGGTAGAT GGTTTCGACG TCGGCGCAAA AACCGGTACG GCGCGTAAGT TGGTTAACGG TCGTTACGTC GATTACAAAC ACGTTGCCAC TTTCATCGGT TTTGCCCCGG CTAAAAATCC GCGGTGATT GTGGCGGTAA CCATTGACGA GCCGACTGAC AACGGTTACT ACGGCGGCGT AGTGACAGGT CTCAAACATCT GGGCGGTAGC TCGAACATCT TCGGCGTTAC TCCGACCAAA CCTCTGACCA AT
length: 402
status: Sanger trace checked
date entered: 2006-09-04
datestamp: 2006-09-04
sender: Muhamed-Kheir Taha, National Reference Center for Meningococci, Institut Pasteur, France curator: Keith Jolley, University of Oxford, UK
mutation FSO4L; yes
mutation ASION: yes
mutation I515V: yes
mutation H541N: yes
mutation I566V: yes
Publication (1)
<ul> <li>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, Skozynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007). Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis. Antimicrob Agents Chemother 51:2784-92</li> </ul>
Isolate databases
PubMLST isolates: Contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample. 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.
--ignore_multiple_hits
    Set allele designation to 'N' if there are multiple designations set for
   a locus. The default is to use the lowest allele value 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).
```

(continues on next page)

(continued from previous page)

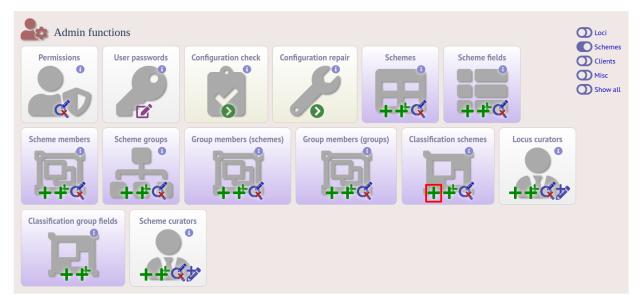
```
--match_missing
   Treat missing loci as specific alleles rather than 'any'. This will
   allow profiles for every isolate that has <= threshold of missing alleles
   to be defined but may result in some isolates having >1 ST.
--max ID
   Maximum isolate id.
--min ID
   Minimum isolate id.
--min_size 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 VTEW
   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.

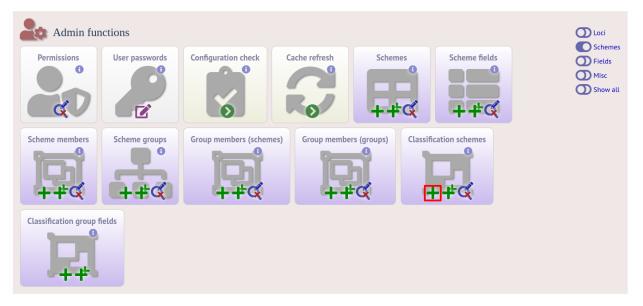
The status can be 'experimental' or 'stable'. The status of a scheme will be shown in the web interface to indicate that any groupings are subject to change and do not form part of the stable nomenclature.

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	<b>†</b>
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: <b>2020-07-19</b>	
<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 scheme	
Add new classification scheme	
Please fill in the fields below - required fields are marked with an exclamation mark (!). Record	+
id:1 2 scheme id:1 N. meningitidis cgMLST v1.0 ∽	
name:! Nm_cgc_25	$\odot$
	wed between profile and at least one group member profile.
use relative threshold: Otrue false Calculate threshold using ratio of shared/pre status: experimental	ent in our prontes in panwise comparison.
curator:! Keith Jolley (keith)	
datestamp:/ 2020-07-19	
description: Single linkage clustering with each group member <u>allelic</u> mismatches to at least one other member are ignored in comparisons.	
seqdef cscheme id: 2 G cscheme_id number defined in seqde	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:

```
{\tt 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*.

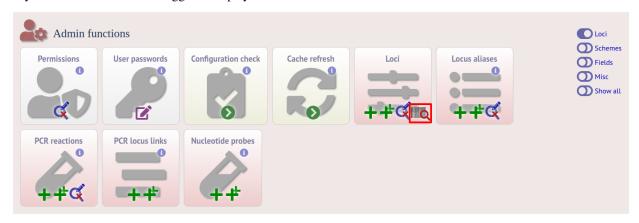
Clicking the hyperlinks will take you to a table containing matching isolates, from where standard analyses can be performed.

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     • =     • Enter value       Display/sort options     Action	+ 3
Display/sort options Action Order by: id  Display: 25  records per page	
4 records returned. Click the hyperlinks for detailed information.         Your projects       Bookmark query         Select project       Add these records       2020-07-19:1	
Isolate fields () MLST Finetyping antigens Id Isolate aliases   country  year   disease   species   capsule group  ST   clonal complex   PorA_VR1   PorA_VR2   FerA_VR	
662 2837 M97/252508 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         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           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 > Organism > Neisseria isolates > Scan EMBL/Genbank record	
Scan EMBL/Genbank record for loci	2
This function allows you to scan an EMBL or Genbank (whole genome) file in order to create a batch upload file for setting up new loci.         Please enter accession number       Primary identifier         Accession:       AM421808         Image: Comparison of the primary identifier       Action         SUBMIT       SUBMIT	
Download table Suitable for batch upload of loci. Download alleles Suitable for defining the first allele in the seqdef database.	
Annotation information accession: AM421808 version: 1 type: dna length: 2194961 description: Neisseria meningitidis serogroup C FAM18 complete genome. coding regions: 1975 Coding sequences	
Locus         Aliases         Product           NMC0001         lpxC;         UDP-3-0-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase	Length 924
envA NMC0002 pilS1 pilin (fragment)	291
NMC0003 pilS2 truncated pilin	366
NMC0004 fbp peptidyl-prolyl cis-trans isomerase NMC0005 putative membrane protein	330 219
NMC0005 putative investment of the second seco	954
NMC0007 metG methionyL-tRNA synthetase	2058
NMC0008 glmS glucosaminefructose-6-phosphate aminotransferase [isomerizing]	1839
NMC0009 putative lipoprotein NMC0010 gna33 outer membrane lipoprotein Gna33 (1)	519 1326
NMC0010 grass dote memorane upprotein Grass () 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 Setting up LINcode definitions for cgMLST schemes

Note: The idea behind LINcodes is described in Hennart et al. 2021 bioRxiv 2021.07.26.453808.

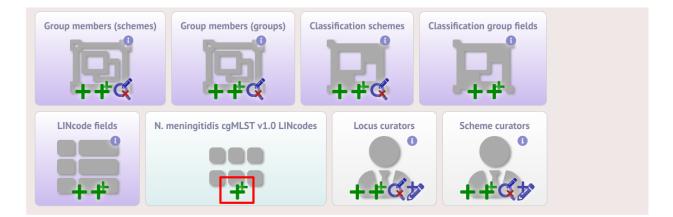
LINcode schemes can be defined by administrators by clicking the add LINcode scheme button within the 'Schemes' group on the curator index page of both the sequence definition and isolate databases. This function is normally hidden, so you may need to click the 'Schemes' toggle to display it. An indexed scheme, e.g. MLST or cgMLST whereby a field defines each unique combination of alleles, needs to be defined before the link will be enabled.



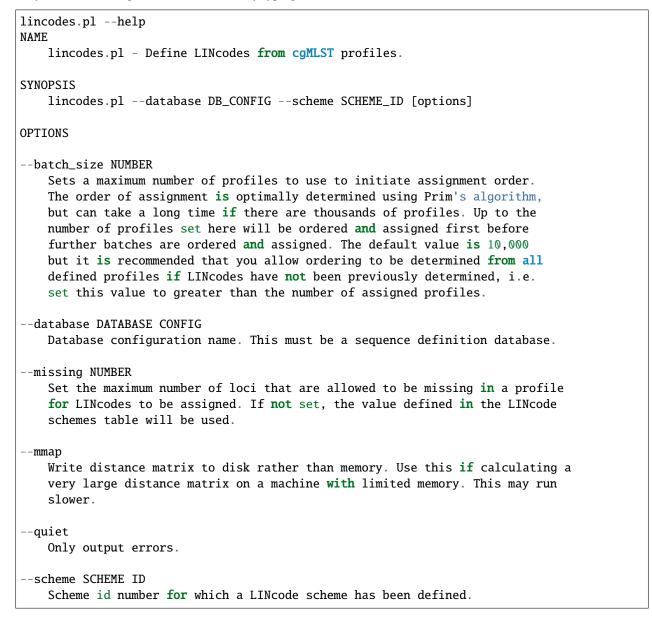
Select the indexed scheme from the dropdown list and enter your locus thresholds in descending order as a semi-colon separated list, e.g. 500;250;100;50;25;10;5;2;1;0. Also enter the number of missing alleles that are allowed within a profile for a LINcode to be assigned. Click 'Submit'.

BIGSdb			
Home > Organisms > 0	rganism > Neisseria typing > Add new LINcode scheme		
Add new I	INcode scheme		User interface
Please fill in the fiel Record	Is below - required fields are marked with an exclamation mark (!).	Action	
scheme id:!	N. meningitidis cgMLST v1.0 v	RESET	SUBMIT
thresholds:	500;250;100;50;25;10;5;2;1;0 Semi-colon separated list of thresholds	RESET	SUBMIT
max missing:!	50 Sumber of loci that are allowed to be missing for a LINcode to be defined.		
curator:	Keith Jolley (keith)		
datestamp:!	2022-03-07		

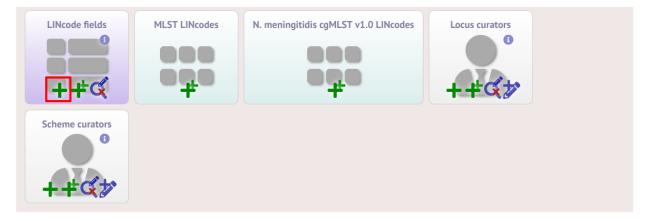
If LINcodes have been previously defined, the existing assignments can be uploaded by clicking the batch add LINcodes button (sequence definition database). This allows you to copy and paste assignments from a spreadsheet template. The template consists of the index field (e.g. cgST), and separate columns for each threshold level.



LINcodes can then be assigned automatically using the lincodes.pl script found within the scripts/maintenance directory. A full list of arguments can be found by typing:



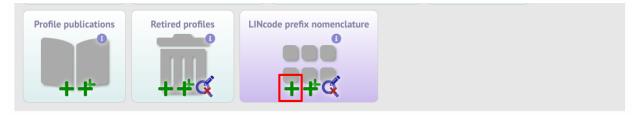
It is also possible to provide nomenclature for specific LINcode prefixes. These can be used to define lineages or sublineages. LINcode fields are first defined by clicking the add lincode field button.



Enter the name of the field you wish to define, e.g. 'lineage' and the data type for any values (either text or integer).

BIGSdb				
Home > Organisms > Org	ganism > Neisseria typing > Add new LINcode field			
Add new L	INcode field			Userinterface
Please fill in the fields	below - required fields are marked with an exclamation mark (!).			
Record		Action		
scheme id:!	N. meningitidis cgMLST v1.0 v	RESET	SUBMIT	
field:	lineage	KESET	SUBMIT	
type:!	text v			
	Keith Jolley (keith)			
	2022-03-08			
display order:	\$			

A new menu item called 'LINcode prefix nomenclature' will appear in the curator part of the index page. It is hidden by default so you may need to click the 'Show all' button. Click the add button.



Enter a LINcode prefix (left-hand part of LINcode) and assign a name to it. Make sure you do the same in both the sequence definition and isolate databaase. You will then be able to search isolates based on these field values and they will appear in isolate records.

BIGSdb					
Home > Organisms >	Organism > Neisseria typing > Add new LINcode prefix nomenclature				
Add new	LINcode prefix nomenclature				User interface
Record scheme id:1 prefix:1 field:1	elds below - required fields are marked with an exclamation mark (!). N. meningitidis cgMLST v1.0 v 1_0_0 lineage v Lineage 3	i	Action RESET	SUBMIT	
curator:! datestamp:!	Keith Jolley (keith) 2022-03-08				

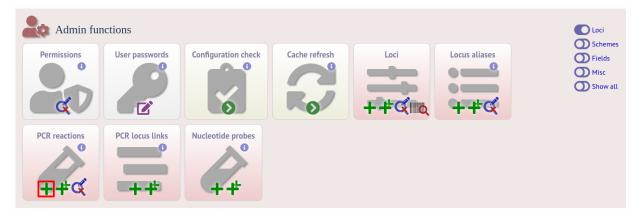
# 5.24 Genome filtering

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.

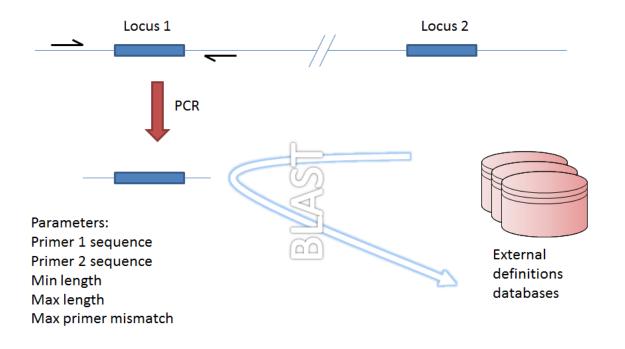
#### 5.24.1 Filtering by in silico PCR

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.

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.



Locus 1 and locus 2 share allele pool

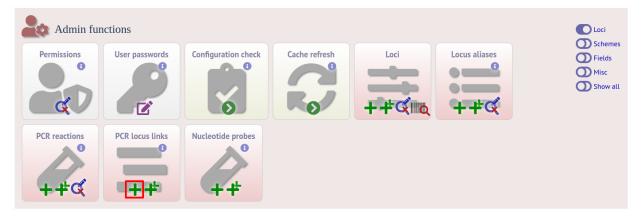
Fig. 1: Genome filtering by in silico PCR.

PubMLST			
Home > Organisms > Organism	> Neisseria isolates > Add new PCR reaction		
Add new PCR r	eaction		
Please fill in the fields below - re	quired fields are marked with an exclamation mark (!).		
Record		Action	
id:!	2	RESET SUBMIT	
description:!		RESET SUBMIT	
primer1:!			
primer2:!			
	Keith Jolley (keith)		
datestamp:!			
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.

### 5.24.2 Filtering by in silico hybridization

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

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

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

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

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

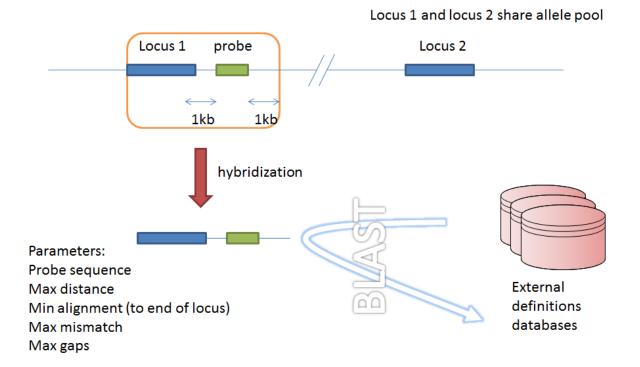


Fig. 2: Filtering by in silico hybridization



- 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.

## 5.25 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 > Organ	ism > Neisseri	ia isolates > Query or u	odate isolates				
Query or upda	ate isol	ates					🛓 Help 🕜 🚯 🎢
Enter search criteria or leav	e blank to brov	wse all records. Modify f	orm parameters to filter o	or enter a list of va	lues.		
	elds			—Display/sort o	ptions		
isolate	<b>∨</b> =	✓ MC58	+ ()	Order by: in Display: 2	1 5 👻 records per page		scending 🗸
Action RESET SEAR	сн						
——Delete —— — Tag s	canning — –	Projects			— Your projects —		
Delete ALL	Scan	Select project		✓ Link	Select project	✓ Add these record	is
Bookmark query							
2020-07-19:1	Add book	kmark					
Delete Update Sequence bir		id isolate alia	ases country ye			e group ST clonal complex Por B 74 ST-32 complex 7	Finetyping antigens VR1 PorA_VR2 FetA_VR 16-2 F1-5

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

E Publications (7)
Bennett JS, Bentley 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: insights gained from the genome of Neisseria lactamica isolate 020-06. BMC Genomics 11:652 78 isolates
Bennett JS, Jolley 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 within the genus Neisseria. Microbiology 158:1570-80 55 isolates
• Bennett JS, Jolley 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: 1
length: 2272360 bp
loci tagged: 2,230
Show sequence bin

Click the 'Renumber' button:

quence	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	←		Rendiniber
					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'.

ubMLST		
ome > Organisms > Organism > Neisseria isolates > Renu	umber locus genome positions	
Renumber locus genome pos	sitions based on tagged sequences	
You have selected to renumber the genome positions set in	n the locus table based on the tagged sequences in sequence id#1.	
Option	Action	
Remove positions for loci not tagged in this sequence		
Chemore positions for toer not tagged in this sequence	RENUMBER	
Locus + Existing genome position + Ne		
NEIS2139	7	
NEIS2140 NEIS2141	502 918	
NEIS2141	2517	
NEIS2142	3158	
NEIS2144	3635	
NEIS2145	4311	
NEIS2146	4958	
NEIS2147	5936	
NEIS2148	6281	
NEIS2149	7573	
tRNA-lys		
	9197	
NEIS2150	9197 9346	
NEIS2150 NEIS2151		

# 5.26 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])

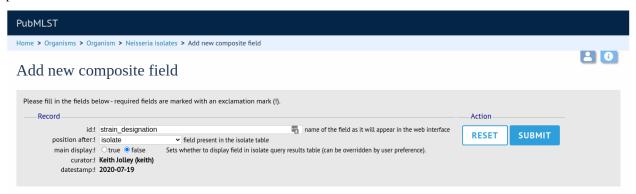
e.g. A: P1.5-2,10: F1-5: ST-4 (cc4)

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

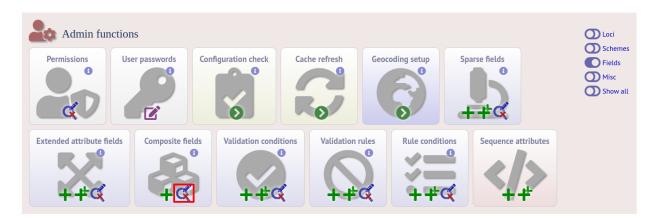
Composite fields can be added to the database by clicking the add (+) composite fields link on the curator's main page. This function is normally hidden, so you may need to click the 'Fields' toggle to display it.



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.



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.	
Delete       Update       field name       position after       missing data         X       strain_designation       isolate       false       [capsule_group]: P1.[PorA_VR1],[PorA_VR2]: [FetA_VR]: ST-[scheme 1:ST] ([scheme 1:ST] ([scheme 1:ST])       ND: P1.ND,ND: F-ND: ST-	ND (-)

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

bMLST						_	_	_	
ne > Organisms > Organism	n > Neissei	ria isolates > Up	date composite field						
pdate compos	site fi	eld - str	ain_desig	nation					
Position/display position after: isolate main display: O true ® fal UPDATE	se	~							
field e	mpty value		regex	curator	datestamp	delete	edit	mov	/e
capsule_group [isolate field]	ND				2020-07-16	×	1		¥
: P1.					2009-11-12	×			↓
PorA_VR1 [locus]	ND				2009-11-12	×		ريني	◆
,					2009-11-12	×			↓
PorA_VR2 [locus]	ND				2009-11-12	×			•
:					2009-11-12	×			•
FetA_VR [locus]	F-ND				2009-11-12	×		<u> </u>	<u>*</u>
: ST-					2009-11-12	×		ريني	<u>*</u>
ST [MLST field]	ND				2009-11-12	×			<b>↓</b>
( lonal_complex [MLST field]		c/ST (\S+)	complex.*/cc\$1/		2009-11-12	×			<b>↓</b>
tonat_comptex [MLS1 netd]		3/31-((3+)	compiex. /ccai/		2009-11-12				
) Add new field:				Rentri Jotley	2009-11-12	<b>^</b>	9		⊻
text field:		!	+						
isolate field:		<b>~</b> +	·						
locus field:			<b>* +</b>						
tocus netu.									

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'.

# 5.27 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.

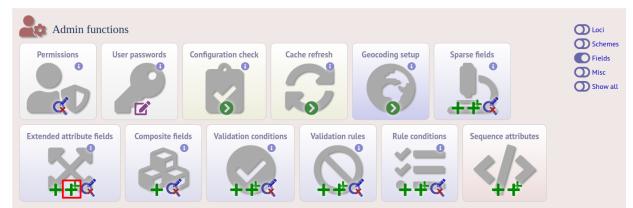


Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

- 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	E	3
Instructions This page allows you to upload isolate field extended attribute 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 The text the description of database fields for help with filling in the template. Upload Paste in tab-delimited text (Include a field header line).	Action	

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

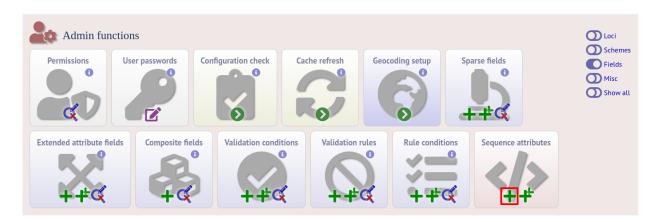
isolate_field	attribute	field_value	value
country	continent	Afghanistan	Asia
country	continent	Albania	Europe
country	continent	Algeria	Africa
country	continent	Andorra	Europe
country	continent	Angola	Africa

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

# 5.28 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		20
Please fill in the fields below - required fields are marked with an exclamation mark (!).  Record  key:: read_length type:: integer > curator:: Keth Jolley (keith) datestamp:: 2020-07-19 description:	Action SUBMIT	ł

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 form <i>Please note that you can reach this page for a specific isolate by querying isolates and th</i> Please fill in the following fields - required fields are marked with an exclamation mark ( Paste in sequences in FASTA format:	hen clicking 'Upload' within the isolate table.	
Attributes isolate id: ! Enter isolate id sender: ! Select sender method: run id: assembly id: read length:	Options         Don't insert sequences shorter than 100 v bps.         Don't insert sequences containing only homopolymers.         Link to experiment:         v         Alternatively upload FASTA file         Choose file         No file chosen         Action         RESET         SUBMIT	

# 5.29 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.

	> Organism > Neisseria isolates	Configura	CON CHECK -	
onfigura	tion check No.	iccori	a icol	
Joinigura	tion check - Ne	isseria	a 1501a	
PostgreSQL				
0 1				
BIGSdb requires Pg	9.5 or higher. You are running vers	sion 10.12.	<ul> <li>A</li> </ul>	
Helper appli	cations			
Program	Path		Executable	
EMBOSS infoalign		~	~	
EMBOSS sixpack EMBOSS stretcher	/usr/bin/sixpack /usr/bin/stretcher	<i></i>	- <u>`</u>	
GrapeTree	/usr/local/GrapeTree/grapetree.py		•	
blastn	/usr/bin/blastn	, <b>.</b>	~	
blastp	/usr/bin/blastp	~	- V	
blastx	/usr/bin/blastx	× .	×	
blat	/usr/local/bin/blat	~	~	
clustalw	/usr/bin/clustalw	~	<ul> <li>Image: A set of the set of the</li></ul>	
ipcress	/usr/bin/ipcress	<ul> <li>Image: A set of the set of the</li></ul>	<ul> <li>Image: A second s</li></ul>	
mafft	/usr/bin/mafft	<ul> <li>Image: A set of the set of the</li></ul>	<ul> <li>Image: A set of the set of the</li></ul>	
makeblastdb	/usr/bin/makeblastdb	<ul> <li>Image: A set of the set of the</li></ul>	<ul> <li>Image: A set of the set of the</li></ul>	
mogrify	/usr/bin/mogrify	<ul> <li>Image: A set of the set of the</li></ul>	<ul> <li>Image: A second s</li></ul>	
muscle	/usr/bin/muscle	× .	× .	
tblastx	/usr/bin/tblastx	~	~	
Locus databa	uses (only showing loci	with po	otential	ems - show all loci)
		-		
Locus	Database	Host	Port Id	ue Database accessible Sequence query Sequences assigned
Locus NEIS0895 (parA)		Host zoo-aberlou	Port Id	ue   Database accessible   Sequence query   Sequences assigned
Locus NEIS0895 (parA) NEIS0903 (opaD)	Database pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou	Port Id Jr 5432 N Jr 5432 N	ue Database accessible Sequence query Sequences assigned 5
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N           ur         5432         N           ur         5432         N           ur         5432         N	ue   Database accessible   Sequence query   Sequences assigned 5
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N	ue     Database accessible     Sequence query     Sequences assigned       5     V     V     X       3     V     X     X       4     V     X     X       5     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N	ue     Database accessible     Sequence query     Sequences assigned       5     V     V     X       3     V     X     X       4     V     X     X       5     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N	ue     Database accessible     Sequence query     Sequences assigned       5     V     V     X       3     V     X     X       4     V     X     X       5     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N	ue     Database accessible     Sequence query     Sequences assigned       5     V     V     X       3     V     X     X       4     V     X     X       5     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013 NEIS2538	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N	ue     Database accessible     Sequence query     Sequences assigned       5     V     V     X       3     V     X     X       4     V     X     X       5     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013 NEIS2538	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N	ue     Database accessible     Sequence query     Sequences assigned       5     V     V     X       3     V     X     X       4     V     X     X       5     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2538 Scheme data	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N	ue     Database accessible     Sequence query     Sequences assigned       5     V     V     X       3     V     X     X       4     V     X     X       5     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X       3     V     X     X
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013 NEIS2538 Scheme data	Database pubmist_bigsdb_neisseria_seqdef pubmist_bigsdb_neisseria_seqdef pubmist_bigsdb_neisseria_seqdef pubmist_bigsdb_neisseria_seqdef pubmist_bigsdb_neisseria_seqdef pubmist_bigsdb_neisseria_seqdef bases description	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N	ue Database accessible   Sequence query   Sequences assigned 5  5  5  5  5  5  5  5  5  5  5  5  5
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013 NEIS2538 Scheme data	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef basses description #LST pubmlst_b igsdb_	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           yr         5432         N	ue     Database accessible     Sequence assigned       5     ✓     ✓       3     ✓     ✓       4     ✓     ✓       5     ✓     ×       4     ✓     ×       5     ✓     ×       4     ✓     ×       5     ✓     ×       6     ✓     ×       7     ×     ×       8     ✓     ×       9     ✓     ×       9     ✓     ×       9     ✓     ×       9     ✓     ×       9     ✓     ×
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013 NEIS2538 Scheme data Scheme data Scheme fata	Database           pubmist_bigsdb_neisseria_seqdef           pubmist_bigsdb_neisseria_seqdef           pubmist_bigsdb_neisseria_seqdef           pubmist_bigsdb_neisseria_seqdef           pubmist_bigsdb_neisseria_seqdef           pubmist_bigsdb_neisseria_seqdef           pubmist_bigsdb_neisseria_seqdef           bases           description           HLST         pubmist_bigsdb_neisseria_seqdef           bmal_MLST         bigsdb_neisseria_seqdef	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ur         5432         N           eria_seqdef         s_seqdef           eria_seqdef         eria_seqdef	ue Database accessible   Sequence query   Sequences assigned 5
Locus NEIS0895 (parA) NEIS0903 (opaD) NEIS1454 NEIS1551 (opaC) NEIS2013 NEIS2538 Scheme data Scheme data Scheme N Riboss rpIF N.meningiti	Database pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef pubmlst_bigsdb_neisseria_seqdef bases description 4LST pubmlst_b species pubmlst_b dis cgMLST v1.0 pubmlst_b	Host zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou zoo-aberlou	Port         Id           ır         5432         N           ır         5432         N           ır         5432         N           ır         5432         N           yr         5432         N	ue     Database accessible     Sequence assigned       5     ✓     ✓       3     ✓     ✓       4     ✓     ✓       5     ✓     ×       4     ✓     ×       5     ✓     ×       4     ✓     ×       5     ✓     ×       6     ✓     ×       7     ×     ×       8     ✓     ×       9     ✓     ×       9     ✓     ×       9     ✓     ×       9     ✓     ×       9     ✓     ×

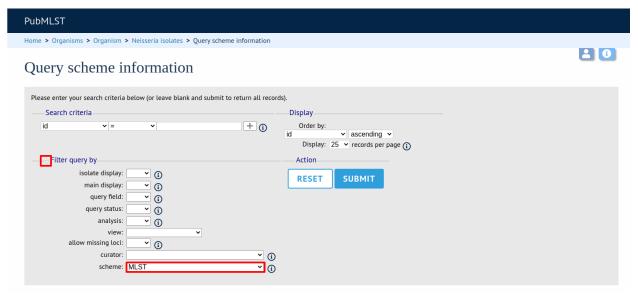
Any problems will be highlighted with a red X.

# 5.30 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 sch	neme information										
Query scheme i	nformation										2	
Please enter your search criteria	below (or leave blank and subm	it to return all record	ds).									
Search criteria			Display									
id 🗸 =	~	+ ()	Order b									
			id Displa	→ ay: 25 →	ascendin records p		)					
			Action			_	•					
isolate display:	× (i)		RESE	T SI	ЈВМІТ							
main display:	U		RESE									
query field:	0											
query status: analysis:												
view:	~ ·											
allow missing loci:	<b>`</b> ()											
curator:		<b>~</b> (i)										
scheme:	MLST	~ (i)										
1 record returned.												
	se configuration—											
	nfiguration/data											
Delete Update id name	dbase name dbas		isolate display*	main display*	query field*	query status*	analysis* vie	ew display order	allow missing loci	curator	datestamp	date entere
🗙 🖋 1 MLST pubmls	t_bigsdb_neisseria_seqdef	1						1		Keith	2019-07-	2009-
Default values are displayed for this	field. These may be overridden by us	er preference								Jolley	26	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
<pre>id description dbase_name dbase_host dbase_port dbase_user dbase_password dbase_ →table isolate_display main_display query_field query_status analysis display_ →order allow_missing_loci curator datestamp date_entered 1 MLST pubmlst_bigsdb_neisseria_seqdef mv_scheme_1 1 1 1 1 1 1 2. → 2012-03-22 2009-11-12</pre>
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
1 gdh 5 2 2009-11-12
1 pdhC 6 2 2009-11-12
1 pgm 7 2 2009-11-12
scheme_fields
<pre>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&amp;db=pubmlst_neisseria_ →seqdef&amp;scheme_id=1&amp;profile_id=[?] 1 1 1 0 2 2010-01-20 1 clonal_complex text 0 2 1 1 1 1 2 2009-11-16</pre>

# 5.31 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.
SYNOPSTS
   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).
```

# 5.32 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.
```

# 5.33 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

SIX

# **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	
Add new user	🛓 Help 🗹 🕄
Please fill in the fields below - required fields are marked with an exclamation mark (l).  Record  id: 548  user name: bloggs first name: bloggs first name: Joe email: joe.bloggs@zoo.ox.ac.uk affiliation: University of 0xford, UK  status: user  date entered: 2020-07-22 datestamp: 2020-07-22 curator: Keith Jolley (keith) submission emails: True filse ① Receive new submission E-mails (curators only)	Action RESET SUBMIT
account request emails: Otrue I 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	

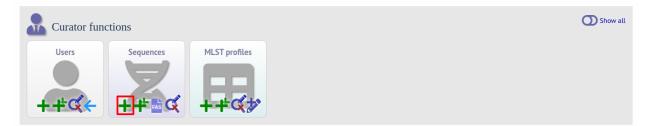
Normally the status should be set as 'user'. Only admins and curators with special permissions can create users with a status of curator or admin.

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	Help 🗹 🕚
Please fill in the fields below- required fields are marked with an exclamation mark (!).  Record  locus:: abc2  ittraction  sequence:: TTTGATACCGTTGCCGAAGGTTTGGGTAAAATTCGCGATTTATTGCGCCGTTACCACCGCGTCATGAGTTGG AAAACGGTTCGGGGTGAGGCTTTGGGCGAAACTCAACGAATAACGAAAAATCGAAGCAGACGGCTGGAA GCTGGGTGCGGCAGTCAGGGATGTGGCGATGTGGGGAAAACTCAACGATTGAAGCAGAAAAATCGGCAACGTGGGAAGCGGCTGGGGAGGATGGGCAGCGGATGGGCAGAGGAGCGGCTGGGGACGAACGGCGGATGGGAAAACTCAACGAAAAATCGGCGATTGCTGGCGAAGCGGCTGGGGAAGACGGCGGGGGGGG	Action RESET SUBMIT
status:     Sanger trace checked        sender:     Jolley, Keith (keith)        curator:     Keith Jolley (keith)        date entered:     2020-07-22       datestamp:     2020-07-22       type allele:     Orue O 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	
PubMed ids:	
Genbank ids:	

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 🗹
<ul> <li>Instructions</li> <li>This page allows you to upload allele sequence data as tab-delimited text or copied from a spreadsheet.</li> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>If the locus uses integer allele ids you can leave the allele id field blank and the next available number will be used</li> <li>the status defines how the sequence was curated. Allowed values are: Sanger trace checked; WGS: manual extract (10 WGS: automatically checked; unchecked.</li> <li>Sequence flags can be added as a semi-colon (2) separated list.</li> <li>Sequence flags can be added as a semi-colon (2) separated list.</li> <li>Templates</li> </ul>	BIGSdb)', 'WGS:		t (BIGSdb), "WGS: visual	ly 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 da 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 cod also ignored. Override sequence similarity check		and no internal st	top codons. Existing sec	quences are
Paste in tab-delimited text ( <b>include a field header line</b> ).	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.

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	💄 Help 📝
Import status Sender: Keith Jolley	
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. Note: valid sequence flags are displayed with a red background not red to	ext.
locus         allele_id         sequence         status         type_allele         sender         curator         date_entered         datestamp         co           abc2         1079         TTTGATACCGTTGCCGAAGG         GCGGATTGTGAACTTGACC         WGS: automated extract (BIGSdb)         2         2         2020-07-22         2020-07-22	mments flags
abcZ         1079         TTTGATACCGTTGCCGAAGG         GCGGATTGTTGAACCTTGACC WGS: automated extract (BIGSdb)         2         2         2020-07-22         2020-07-22           abcZ         1080         TTTGATACCCTTGCCGAAGG         GCGGATTGTTGAACCTTGACC WGS: automated extract (BIGSdb)         2         2         2         2020-07-22         2020-07-22	

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	Help 🗹
Problem(s)         Primary key       Problem(s)         locus: abcZ; allele_id: 1079       Sequence contains non nucleotide (AIC[G T) 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 flags         locus lallele_id       sequence         abc2       1079         TTTGATACCGTTGCCGAAGG       CGCGGATTGTTGAACTTGACC         wbcZ       1080         TTTGATACCCTTGCCGAAGG       CGCGGATTGTTGAACTTGACC         wbcZ       1080	sare displayed with a red background not red text.       sender (curator date_entered datestamp comments flags       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		O Show all
Users	Sequences	MLST profiles	

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

PubMLST	
Home > 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 intege ids only). Do not include the locus name in the identifier in the FASTA file.	er
Please note that you can not use this page to upload sequences for loci with extended attributes.  Enter parameters	
locus: labcZ 🗸	
status: Sanger trace checked v	
sender: Jolley, Keith (keith)	
sequence >isolate1 (FASTA): TTTGATACCGTTGCGAAGGTTCCGGCGAAATTGGTGATTTATTGCGCCGTTATCATCAT GTCAGCCATGAGTTGGAAAATGGTTCGAGTGAGGCTTTGTTGAAAGAACTCCAACGAATTG CAACTTGAAATCGAACGCGAAGATGGATGCGGCAGTCGGACACTAGCGACACTTGG GGGGAACTCGGTTTGCCGGAAAATGGAAAAATCGGCAGCCTTCGTCGGCGGTCAGAAAAAG CGCGCGTCGCCTTGGCTGCGAAAATGGACAAAATCGGCAGCCGTCGGACAGCGGCGGTCAGCAAGAAAG CGCGCGTCGCCTTGGGTGCGAAAAATCGGCCACCGACGTATTGCTGCTGGACGACATGGATACGCCGG ACCAACCATTTGGATATCGACGCGCCGCCGTTTTTTGGGCAAAATCTGCTCCAGGCAGCTGGATGGGTGGTGGATACCGACGGCGGTCAGAAGCGCCGGCGGTCGAAAATCGGCCAGCCGGCGT GGCGAACTCGATC GGCGACTCGGTTGGGTTACCCACGGCCGCCGTTTTTTGGGCAAATATCGCCACGCGGGGAT GTCGAACTCGATC >isolate2	
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.	
C Use next available if (only for loci with integer ids)	
Action	
RESET CHECK	

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	etp 🕜 🕄
Sequence check Locus: abcZ Original designation Allele id Status isolate1 1079 OK isolate2 1080 OK UPLOAD VALID SEQUENCES	

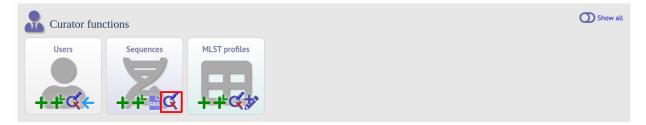
Any invalid sequences will be indicated in this confirmation page and these will not be uploaded (you can still upload the others), e.g.

PubMLST		
Home > Organisms > Organism > Neisseria typing > Batch insert sequences		
Batch insert sequences		Letp 🔀 🛛
Sequence check Locus: abcZ		
Original designation Allele id Status	Action	
isolate1 1079 OK isolate2 1080 Sequence contains 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 attrib	ıte search	Help 🔀 🚺
export plugins. Also note that some loci in this search to a locus that uses integ	5 . 5	rrch for these at the locus-specific query page. Use this page also for access to the sequence analysis or es using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your (s).
Search criteria		Display
allele id 🔹 👻 =	<ul> <li>✓4</li> <li>(+) ()</li> </ul>	Order by: locus ascending Display: 25 records per page (
Filter query by locus: type allele: sender: curator: allele flag:	abcZ • (i)	Action RESET SUBMIT

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:		Š
	Sanger trace checked	
type allele:		
	Keith Jolley	
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.

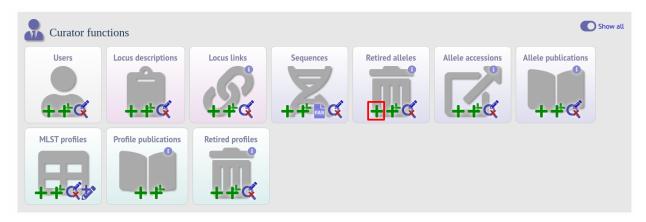
Note: Adding flags and comments to an allele record requires that this feature is enabled in the *database configuration*.

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: 1 sequence: l TTTGATACTG TTGCGCGAAGG TTTGGGCGAA ATTCGCGATT TATTGCGCCG TTATCATCAT GTCABCCATG AGTTGGAAA TGGTTCGAGT GAGGCCTTAT TGAAAGAGGCT CAACGAATTG CAACTTGAGA TCGAACGAA GGACGGCTGG AAGTTGGGATC CGGCGGTGAA CCAACTTG	RESET	Å
status:       Sanger trace checked          sender:       Jolley, Keith (keith)          curator:       Keith Jolley (keith)          date entered:       2001-02-07         dates entered:       2002-07-72         type allele:       True O 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	<b>2</b> 0
Please fill in the fields below - required fields are marked with an exclamation mark (!).  Record Action Action Action Curator:! NEISO844 Curator:! Keith Jolley (keith) datestamp:! 2020-07-22	ť

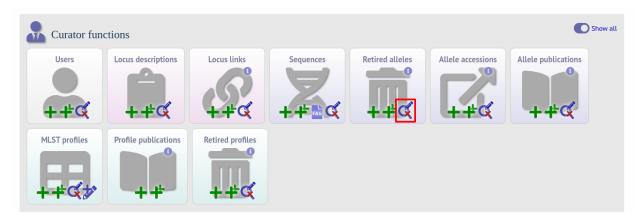
You cannot retire an allele that already exists, so you must delete it before retiring it. Once an identifier is retired, you will not be able to create a new allele with that name.

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).	
Search criteria Display	
datestamp <ul> <li>itoday</li> <li>itoday</li></ul>	
Filter query by Action	
locus: abcZ · G RESET SEARCH	
1 record returned.	
Delete	
Delete ALL	
Delete     Update     locus     allele id     curator     datestamp            × $abcZ$ 1079         Keith Jolley         2020-08-05 $abcZ$ $bbcZ$ $abcZ$ $bbcZ$ $abcZ$ $abcZ$ $bbcZ$ $abcZ$	

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	
datestamp v = v today + i Order by:	
locus	
Filter query by Action	
locus: abcZ  Curator:  Cur	
1 record returned.	
Delete	
Delete ALL	
Delete     Update     locus     allele     id     curator     datestamp       X     Image: Above the stamp     Image: Above the stamp     Image: Above the stamp     Image: Above the stamp       X     Image: Above the stamp     Image: Above the stamp     Image: Above the stamp     Image: Above the stamp       X     Image: Above the stamp     Image: Above the stamp     Image: Above the stamp     Image: Above the stamp	

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Delete retired allele id	
Delete retired allele id	2
You have chosen to delete the following record. locus: abcZ allele id: 1079 curator: Keith Jolley datestamp: 2020-08-05 MELETE	

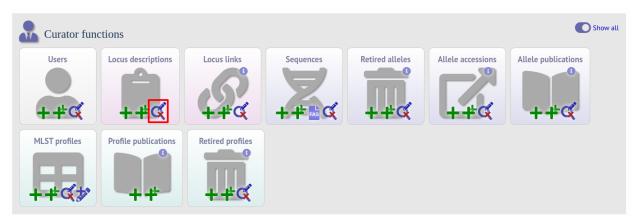
The identifier can now be re-assigned when adding a new allele.

## 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 information	
Query locus description information	20
Please enter your search criteria below (or leave blank and submit to return all records).	
Search criteria Display	
locus     Image: state	
Control Contr	

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).	

Search criteria	Display
	Order by:
	locus v ascending v Display: 25 v records per page (j
ilter query by	Action
locus: NEISO620 (maeA)	RESET SUBMIT
curator: v j	

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     v     =     v     +     •       Iocus     v     ascending     v       Display:     25     v     records per page ()	
Filter query by     Iocus: NEISO620 (maeA)     Curator:     curator:     common name:     ①	
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	n > Neisseria typing > Update locus description		
Update locus d	escription		
Record	required fields are marked with an exclamation mark (!).  NEIS0620 Keith Jolley (keith) 2020-07-22	Action RESET SUBMIT	
full name:			
product:	malate oxidoreductase (EC 1.1.1.38)		
description:	Final step in TCA cycle producing oxaloacetate.		
aliases:	NG00240 ÷		
PubMed ids:	14917678		
links: (Format: URL description)	http://www.enzyme-database.org/ <u>guery.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 () 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: ! 1222 gdh: ! 8 pdhC: ! 4 pgm: ! 6 sender: ! Jolley, Keith (keith) clonal_complex: clonal_complex: curator: ! Keith Jolley (keith) date_entered: ! 2020-07-22 datestamp: ! 2020-07-22 PubMed ids:  Action  RESET 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.

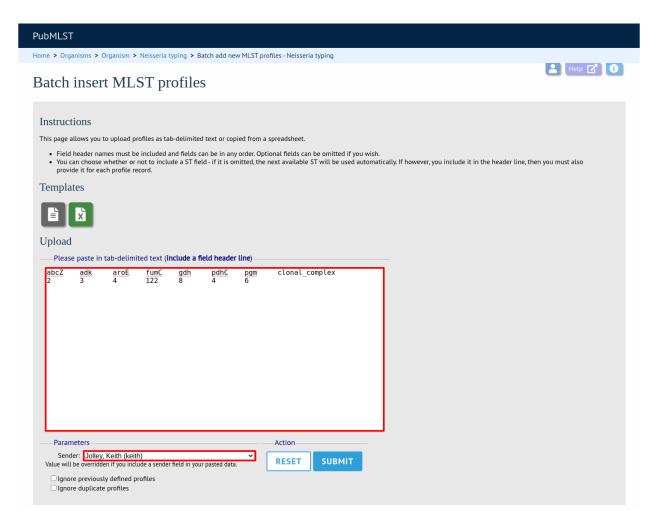
Click the batch add (++) profiles link next to the scheme name:

Curator f	functions		O Show all
Users	Sequences ++	MLST profiles	

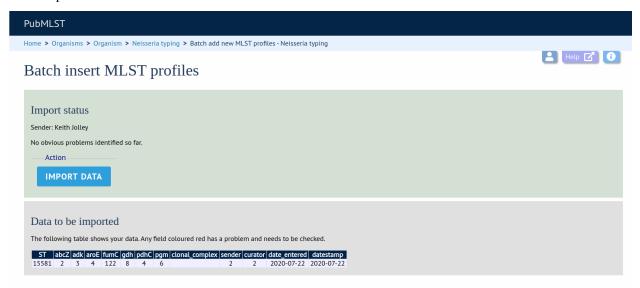
Click the 'Download submission template (xlsx format)' link to download an Excel submission template.

PubMLST	
Home > Organisms > Organism > Neisseria typing > Batch add new MLST profiles - Neisseria typing	
Batch insert MLST profiles	💄 Help 🗹 🚯
Instructions	
This page allows you to upload profiles as tab-delimited text or copied from a spreadsheet.	
<ul> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>You can choose whether or not to include a ST field - if it is omitted, the next available ST will be used automatically. If however, you include it in the header line, ther provide it for each profile record.</li> </ul>	ı you must also
Templates	
Upload	
Please paste in tab-delimited text ( <b>include a field header line</b> )	
Parameters Action	
Sender: Select sender  Value will be overridden if you include a sender field in your pasted data.	
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		O Show a
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	
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  Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.  Display/sort options  Order by: ST  V  ascending V  Display: 25 V  records per page ①	Action RESET SUBMIT
1 record returned. — Delete — Delete ALL	
Delete     Update     ST     JabcZ     adk     aroE     fumC     gdh     pdhC     pgm     clonal complex       X     X     4563     2     7     6     13     9     18     8     ST-167 complex	
Analysis tools	

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 the	e following record. Select 'Delete and Retire' to prevent the identifier being reused.	
abcZ : adk : arcE : fumC : gdh : pdhC : pgm : clonal_complex : sender: curator: date entered:	4563 2 7 6 6 13 9 9 18 8 8 5T-167 complex Ana-Belen Ibarz-Pavon Keith Jolley	
Action		
	TE AND RETIRE	

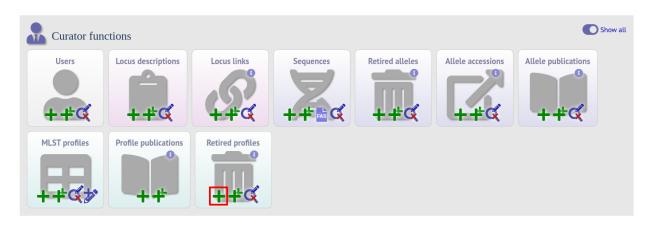
To modify the profile, click the 'Update' link next to the profile following the query. A form will be displayed - make any changes and then click 'Update'.

PubMLST			
Home > Organisms > Or	ganism > Neisseria typing > Update profile		
Update prof	ile		🛓 Help 才 🕄
Record		Action	
ST: abC2: adk: aroE: fumC: gdh: pdhC: pgm: clonal_complex sender: curator: date_entered:	1 7 1 6 1 13 1 9 1 18 1 8 2 ST-167 complex 1 Ibarz-Pavon, Ana-Belen (aibarz) ✓ 1 Ibarz-Pavon, Ana-Belen (aibarz) ✓ 1 Keith Jolley (keith) 2 2005-03-03 1 2020-07-22	RESET SUBMIT	

#### 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:! 57232 curator:! Keith Jolley (keith) datestamp:! 2020-07-22	Action SUBMIT	ľ

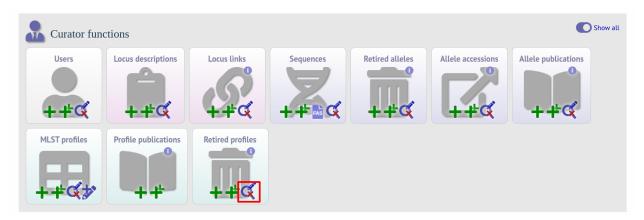
You cannot retire a profile identifier that already exists, so you must delete it before retiring it. Once an identifier is retired, you will not be able to create a new profile with that name.

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.



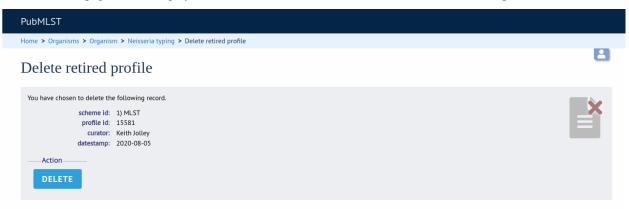
Search by any criteria to find the profile identifier.

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 <ul> <li>icoday</li> <lic< td=""><td></td></lic<></ul>	
Filter query by scheme: MLST  curator:  Scheme:  MLST  Scheme:  S	
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	
Please enter your search criteria below (or leave blank and submit to return all records).	
Search criteria Display	
datestamp v = v today + 0 Vrder by: scheme id v ascending v	
Display: 25 ✓ records per page (j	
Filter query by Action	
scheme: MLST 🗸 🕑 RESET SEARCH	
curator:	
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	Leto 🔽 🕄
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: Jolley, Keith (keith) curator: Keith Jolley (keith) date entered: 2020-07-23 () datestamp: 2020-07-23 () Enter one per line	
Supports multiple values - enter each one on separate line	
year: 2014 () date sampled: dd/mm/yyyy 🗖 ()	
date sampled: dd/mm/yyyy 🗖 👔 date received: dd/mm/yyyy 🗖 👔	
non culture: Otrue Ofalse 👔	
epidemiological year:	
age yr:	
age range:	
age mth:	
sex: V	

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.

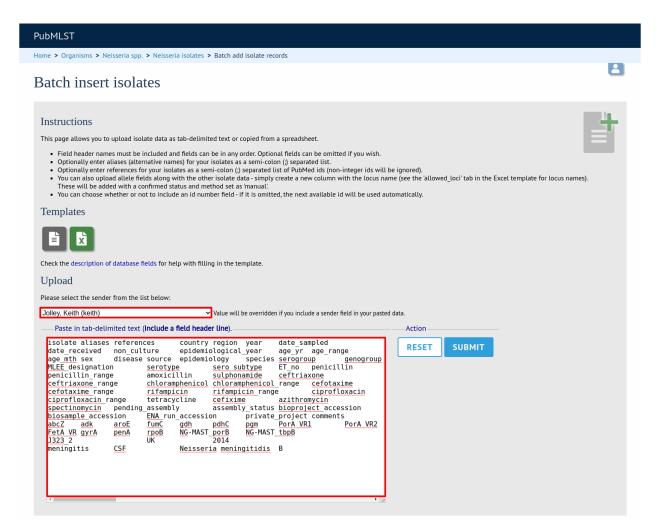
Curator funct	tions			O Show all
Users	Isolates	Sequence bin	Sequence tags	

Download a submission template in Excel format from the link.

PubMLST
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add isolate records
Batch insert isolates
Instructions       If is a get allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.       If is a get allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.       If is a get allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.       If is a get allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.       If is a get allows you to upload isolate as a semi-colon (i) separated list.       If is a get allow for your isolates as a semi-colon (i) separated list of PubMed ids (non-integer ids will be ignored).       If is a get allow for your isolates as a semi-colon (i) separated list of PubMed ids (non-integer ids will be ignored).       If is a get allow for your isolates as a semi-colon (i) separated list of PubMed ids (non-integer ids will be ignored).       If is a get allow for your isolates as a semi-colon (i) separated list of PubMed ids (non-integer ids will be ignored).       If is a get allow for your isolates as a semi-colon (i) separated list of PubMed ids (non-integer ids will be ignored).       If is a get allow for your isolates as a semi-colon (i) separated list of PubMed ids (non-integer ids will be ignored).       If is a get allow for your isolates as a semi-colon (i) separated list of PubMed ids (non-integer ids will be ignored).       If is a get allow for your isolates as a semi-colon (i) separated ids (non-integer ids will be ignored).       If is a get allow for your isolates as a semi-colon (i) separated ids (non-integer ids will be used automatically.       If is a get allow for your isolates as a get allow defined ids (non-integer ids will be used automatically.       If is a get allow for your isolates allow defined ids (non-integer ids will be used automatically.
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). Action           RESET         SUBMIT

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

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



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	8
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 allases references country region year   date_sampled   date_received   non_culture   epidemiological_year   age_yr   age_range   age_mth   sex   disease   source   epidemiological_year   age_yr   age_mth   sex   disease   source   age_mth   sex   disease	gy 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 epidemiological_year age_yr age_range age_mth isex disease isource epidemiology species isorogroup genogroup genogroup_notes 2014 CSF Neisseria B mengitidis	s MLE

## 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 > N	eisseria spp. > Nei	sseria isolates	> Query or upd	ate isolates									
Query or up	odate isol	ates								E	Help		
Enter search criteria or	leave blank to brow	vse all records.	Modify form pa	rameters to	filter	or enter a list of values	5.						
	ce fields					Display/sort optic	ons						
Combine with: AND	~					Order by: id				✓ asce	ending 🗸		
country	~ =	✓ USA		-+ (	i	Display: 25	<ul> <li>records per page</li> </ul>	· (i)					
year	← =	✓ 2009				Action							
154 records returned (:	L - 25 displayed).					RESET	ARCH						
Delete	Tag scanning — –	Projects					-Bookmark guer	v					
Delete ALL	Scan	Select project	t			✓ Link 2	020-07-23:1	, A	dd bookmai	k			
	2 3 4 5 6	00	<b>`</b>	lsc	olate fi	elds j			٢	ILST	Fine	typing anti	gens
Delete Update bin	version id	isolate	aliases	country	year	disease	species	capsule group	ST	clonal complex	PorA_VR1	PorA_VR2	FetA_VR
× 🖋 土	+ 1267	4 M18700		USA 2	2009	invasive	Neisseria	В	7301	ST-32	+	+	+
× / 1	+ 1267	5 M18701		USA 2	2009	(unspecified/other) invasive (unspecified/other)	meningitidis Neisseria meningitidis	В	7302	complex	+	+	+
× 🧪 1	+ 1267	6 M18725		USA 2	2009	invasive (unspecified/other)	Neisseria meningitidis	В	7303	ST-1572 complex	+	+	+
× / ±	+ 1309	0 M19024	PA09015	USA 2	2009	meningitis	Neisseria meningitidis	В	7575		+	+	+

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

PubMLST				
Home > Organisms > Neisseria s	spp. > Neisseria isolates > Delete isolate			
Delete isolate				Help 🗹
You have chosen to delete the formed of the	following record. Select 'Delete and Retire' to prevent the Isola nary metadata	te id being reused.		×
(c country: U continent: N region: W year: 2 Action	M18725 source B: P1.ND,ND: F-ND: ST-7303 species (cc1572) serogroup USA capsule group North America sender WI	: Neisseria meningitidis : B	date entered:	Julia Bennett, University of Oxford, UK (E-mail: julia.bennett@zoo.ox.ac.uk) 2009-03-24 2009-03-24
DELETE DELETE				

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	
Undete isolete	Help 🕜 🚯
Update isolate	
Primary metadata	
id:! <b>12676</b> (j)	
isolate: M18725	
country:! USA 🛛 🗸 🕤	
species: Neisseria meningitidis	
sender:! Zhao, Xin (xinzhao) 🗸 🕤	
curator: Keith Jolley (keith) 👔	
date entered: 2009-03-24 🕢	
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: v ()	
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.

pubs clubs per week: persons kissed past week: regular partner: true fal regular partner smokes: true fal		
ethnicity:		
cumery.		
Loci		
🛓 🛺 All loci		
🔄 🚚 Typing	MLST	
- 💭 MLST - 🎧 eMLST (20 locus partial genes)	abcZ         adk         aroE         fumC         gdh         pdhC         pgm         ST         clon           8         105         52         4         9         11         7         7         7303         ST-15	al complex 72 complex
ocus: '16S_rDNA	✓ 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           Add/update

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.

Home > Organisms > Neisseria spp. > Neisseria isolates > Update allele	
Update abcZ allele for isolate 12676	ļ
♦ Provenance/primary metadata Id: 12676 isolate: M18725 strain designation: B: P1.ND,ND: F-ND: ST-7303 (cc1572) country: USA region: Wi sectes: Invasive (unspecified/other) source: CSF species: Neisseria meningitidis sectes: Neisseria meningitidis sectes: Neisseria meningitidis sectes: Ni. Zhao curator: Julia Bennett coopy-05-24 Curator: Julia Bennett Subory-05-24 Atd/update Existing designations Existing designations	

#### 6.13 Batch updating multiple isolate records

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

Curator fund	ctions			O Show all
Users	Isolates + + C	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	В
101	serogroup	В

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		
secondary fields are unique), this should be entered in The next column should contain the field/locus name id field value 2 country USA 2 abcZ 5 • The columns should be separated by tabs. Any other c • If you wish to blank a field, enter ' <blank>' as the value</blank>	ored. Ed that you are selecting isolates on). If a secondary selection fie the second column. and then the final column should contain the value to be entere olumns will be ignored. e.	
Please paste in your data below:	Options	Allele designations
id field value 100 <u>serogroup</u> B 101 <u>serogroup</u> B	Primary selection field: id  Optional selection field: <a href="mailto:none">none</a> Cupdate existing values  Action  RESET SUBMIT	Add additional new designation

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 > Neisse	ria spp. > Neisseria	a isolates > Batch Isolat	e Update				
Batch isolate	update						L Help
The following changes will Transaction id Field		abase. Please check that (s) currently in database	this is what you intend and th Action	en press 'Update'. If you do	not wish to make these o	changes, press your brows	ser's back button.
1 100 serogrou			update field with new value				
2 101 serogrou	p B	C	update field with new value				
Action UPDATE							

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

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

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

## 6.14 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 update/delete isolates link:



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

PubMLST			
Home > Organisms > Neisseria spp. > Ne	eisseria isolates > Query or update isolates		
Query or update isol	lates		🛓 Help 🔀 🚯 🏏
Enter search criteria or leave blank to bro	owse all records. Modify form parameters to t	filter or enter a list of values.	
Isolate provenance fields		Display/sort options	
Combine with: AND ~		Order by: id	✓ ascending ✓
date entered 🗸 =	✓ 2014-03-18		•
curator (surname) 🗸 =	<ul> <li>✓ Jolley</li> </ul>	Action	
		RESET SEARCH	
3 records returned.			
——Delete —— — Tag scanning —	Projects	Bookmark query	
Delete ALL Scan	Select project	✓ Link 2020-07-23:1 Add book	mark
Delete Update Sequence bin New version		fields () Finetyping antiger	
× / +	id isolate aliases country year dise 28787 M22296 USA	ease species capsule group PorA_VR1 PorA_VR2 Fr Neisseria meningitidis 5-1 2-81	TetA_VR
× / ± +	28787 M22296 USA 28788 M22553 USA	Neisseria meningitidis 5-1 2-81 2-81 2-81 2-81 2-81 2-81 2-81 2-	Ŧ
× 🖉 主 🕂	28789 M22568 USA	Neisseria meningitidis 7-2 🧳 4-39 🗳	+

You will have a final chance to change your mind:

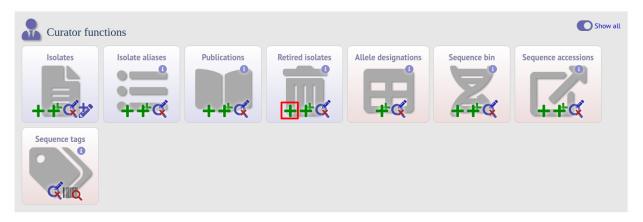
PubMLST			
Home > Organisms	s > Neisseria spp. > Neisseria isolates > Delete multiple records		
Delete m	ultiple isolate records		8
Warning		Confirm action	
	If you proceed, you will delete $\boldsymbol{3}$ isolate records. Please confirm that this is your intention.	DELETE DELETE AND RETIRE	
	The identifiers 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 id	
Add new retired isolate id	
Please fill in the fields below - required fields are marked with an exclamation mark (!).  Record  isolate id: 46262  curator:! Keith Jolley (keith) datestamp:! 2020-07-23	Ħ

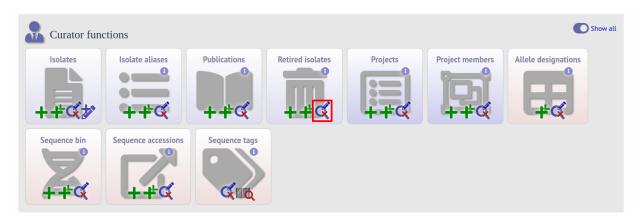
You cannot retire an isolate identifier that already exists, so you must delete it before retiring it. Once an identifier is retired, you will not be able to create a new isolate record using that identifier.

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.

Home > Organisms > Neisseria spp. > Neisseria isolates > Query retired isolate id information Query retired isolate id information	20
Please enter your search criteria below (or leave blank and submit to return all records)Search criteriaDisplay	
datestamp     Image: state is a second in the image: state is a seco	
1 record returned.          Delete       Update       isolate id       curator       datestamp         X       Ø316       Keith Jolley 2020-08-05	

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 <ul> <li></li></ul>	
Filter query by Action Curator: Jolley, Keith (keith) 🕜 🕜 RESET SEARCH	
1 record returned.          Detete Update isolate id curator datestamp         X       80316         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.

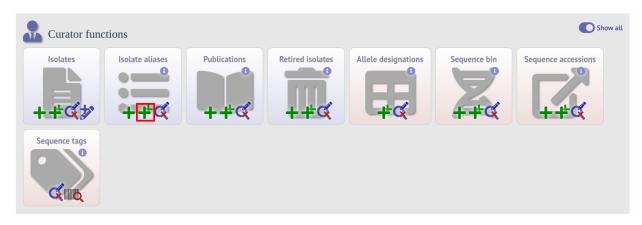
## 6.17 Setting alternative names for isolates (aliases)

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.

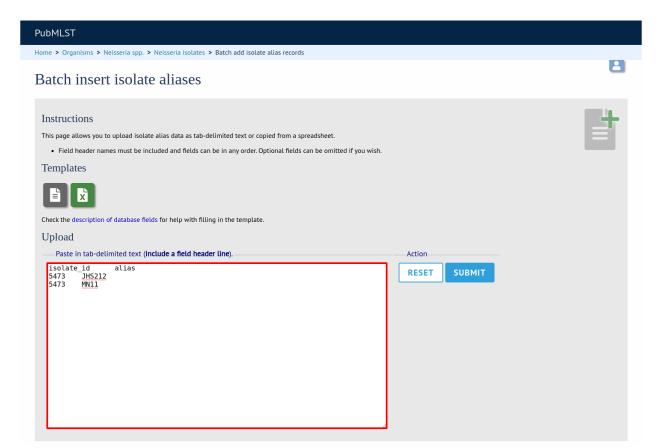
If adding isolates singly, add the aliases in to the aliases box (one alias per line):

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       allas       datestamp       curator         \$473       JHS212       2020-07-23       2         \$473       MN11       2020-07-23       2	

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. <b>Demplates</b> Image: Im	Action RESET SUBMIT	

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		2
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. <b>Demplates</b> Image: The second stable of the second stabl	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 TGTCCACCTTCTTTAATAGAAGTAGTACCAATAACTTTTCCTCACATCGACAGTCCGCAATA TATTGGCCATCACGAGTCATCGATACGGGAGAAGAACTACTTTATTACTTTGAGTATA ACCTTCAATTCGATGACGGAGAAGACGGTAATACCGGTTATCCGCAATAGGCCTATGG AAGTGCCCCTCAAGAACATGCTCAAAGCCAGCAGAACCGGTTATTCGCCATAGGCCATGG AAGTGCCCCCTCAAGAACATGCCTCAAAGCCAGCAGTACCGGTTGTCCCCATCGTTTGCCCATTT GGTGTATATCGAGCAGCCCCCCAACGGCGGTGCCCCCGGTGCCTCCCCATCGG GCCGCCTTCGGACTGCCCCCACAGCGGCTCCCCAACGGTGGCCCACGCGTACCG GCCGCCTTCGGCTTCGACGCCCCCCACAGCGGCTGCCATCCGCTACCGTTACC GCCACCTGGCGCATCCCGCGCCCGCTGGGCCGACGCCTGAGTGCCATCACCGTACCG GCCACCTGGCGCATAAACCCCATAGCTGATTTTACCACCGCGGCCTGGATCGCCATCA GGGATATGCCATGCAGAGCCCCGCTTGGCGAGCCGTGAGTGCCATCGCCACGCG GCCACCTGCAGCATAACCCCCATGGGGTACTTCCAAGGCGCTGGGCCCACGCCTG GGGTATATGCTTATCGCATGGCGATATCTCCCACGGGCCTGGGCCGCCATTA CGGATATGCCTATGCATGGGGTACTTGCAAAATTCAAGGCGCTTGGCCCCCCCC	
Attributes       Options         isolate id: !       3         sender: !       Jolley, Keith (keith)         method:       Illumina         run id:       1         assembly id:       Alternatively upload FASTA file         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	

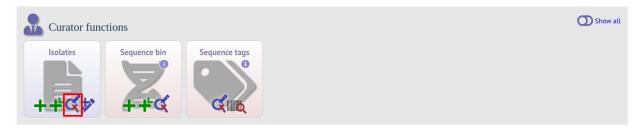
#### Upload sequences

he following sequen	ces will be e	entered.	Summary		Action
inal designation Sequ	ence length	Comments			UPLOAD
180426		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 en	er a list of values.
Isolate provenance fields	splay/sort options
isolate v = v M00242905 + i	Order by: id v ascending v Display: 25 v records per page (i)
Action SEARCH	
1 record returned.	
— Tag scanning — Bookmark query — — — — — — — — — — — — — — — — — — —	
Scan 2020-07-23:1 Add bookmark	
Solution (Located Sequence New Isolate field	lds 🕥 MLST Finetyping antigens
Delete Update bin version id isolate aliases country year dise	se species capsule ST clonal complex PorA_VR1 PorA_VR2 FetA_VR
🗙 🖋 🏦 🕂 3 M00242905 UK 2000 invas (unspecifie	ive 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 Options Options Options Options Don't insert sequences shorter than 100 v bps. 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

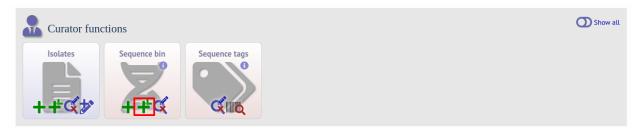
#### 6.19.3 Upload options

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.

Attributes		Options	
isolate id: ! 3) M00242905 sender: ! Jolley, Keith (keith) method: run id: assembly id:	▼ L	elect FASTA file:	

# 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.
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.
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       isolate       current sequence bin state contigs       upload status         80509       JB_21292       -       JB_21292.fasta       X         80510       JB_21293       -       JB_21293.fasta       X         80511       JB_21294       -       JB_21294.fasta       X         Remove       3       FASTA files left to upload.       K	¥
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 y upload multiple times so that the total size of the upload can be larger.	ou can
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				Detek add serve											
Home > Org	ganisms > Neisseri	a spp. > Ne	eisseria isolates >	Batch add new	sequences									А	Ð
Batch	upload s	eauer	nce asser	nblies t	o mul	ltiple	e is	sola	te r	eco	rds				
	-r	- 1													
remove row Remove All files up Actio	Id         Isolate           80309         JB_21292           80310         JB_21293           80311         JB_21294           solate         Solate	current se contigs	equence bin state total size (bp) - - - -	filename IB_21292.fasta IB_21293.fasta IB_21294.fasta	upload statu		ey can I	be adde	ed to th	ne datab	ase.			Z	Z

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         filename         valid FASTA         contigs         total size           80309         JB_21292         B_21292,fasta         ✓         364         2,069,108           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.         ✓         199         2,057,385	
Please do not refresh the page after you click the upload button. Upload may take a short while.	
Attributes Options	
sender: ! Jolley, Keith (keith) 🗸 🗹 Don't in	sert sequences shorter than 100 👻 bps. 🚯
method: Illumina 🗸 🗸 🗸 🗸 🗸 🗸 🗸 🗸 🗸 🗸 🗸 method:	sert sequences containing only homopolymers. 👔
Action	
UPLOAD VALIDATED CONTIGS	

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	<b>2 0</b>
Validation       Id     Isolate     filename     valid FASTA     contigs     total size       80309     JB_21292     JB_21292/fasta     364     2,069,108	
80310 JB_21293 JB_21295.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     ✓       Out't insert sequences containing only homopolymers. ()	
Action	
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					Show all
Users	Isolates + # C *	Sequence bin	Sequence tags			

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	2
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 I Isolate record URI: 1 http://rest.pubmlst.org/db/pubmlst_rmlst_isolates/isolates/933 RESET SUBMIT	

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	
Action 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 t	
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.

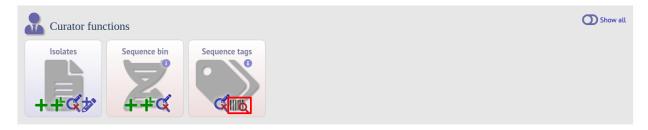
PubMLST	
Home > Organisms > Organism > Leptospira isolates > Add remote contigs	
Add remote contigs	2
Processed contigs Contigs processed: 25 Total length: 2,697,907 bp	

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

#### 6.22 Automated web-based sequence tagging

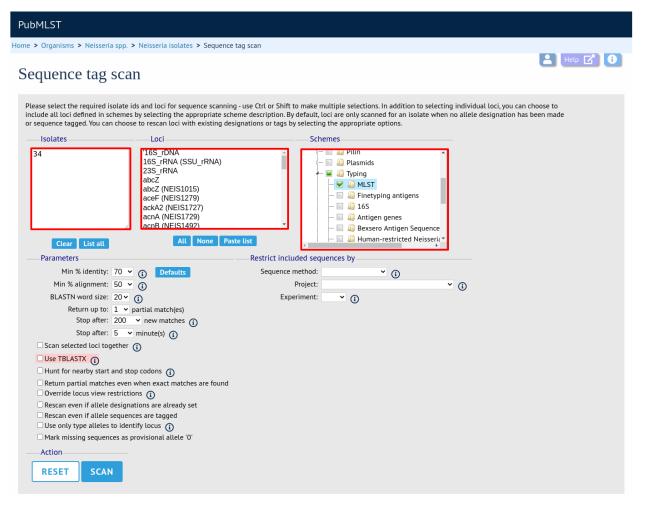
Sequence tagging, or tag-scanning, is the process of identifying alleles by scanning the sequence bin linked to an isolate record. Defined loci can either have a single reference sequence, that is defined in the locus table, or they can be linked to an external database that contains the sequences for 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 V
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 (
<ul> <li>Return partial matches even when exact matches are found</li> <li>Override locus view restrictions (i)</li> </ul>
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.

	Pul	bМ	LST
--	-----	----	-----

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 🖸	+			v
34) 20	exact	adk	3	100.00	465	465	0.0	182750	1392	1856	1392	1856 extract 🖸	<b>→</b>			```
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 🕄	<b>→</b>			<b>`</b>
34) 20	exact	gdh	1	100.00	501	501	0.0	182852	7516	8016	7516	8016 extract 🕥	<b>→</b>			
34) 20	exact	pdhC	1	100.00	480	480	0.0	182819	13868	14347	13868	14347 extract 🕄	<b>→</b>			
34) 20	exact	pgm	3	100.00	450	450	0.0	182831	24559	25008	24559	25008 extract 🔊	<b>→</b>			v
Act TA		.ELE	S/SEQ	DENC	ES									All None	All None	
nished apsed	: Thu Ju time: 5	l 23 1 second		2020	in on the ser	ver for 7 (	days.									

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

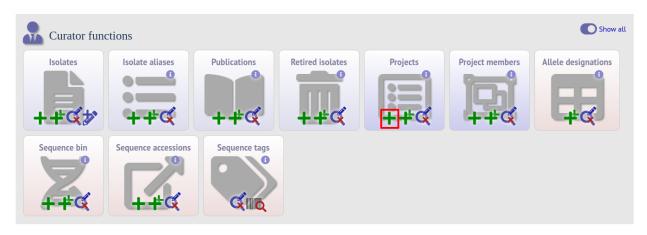
# 6.23 Projects

#### 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:

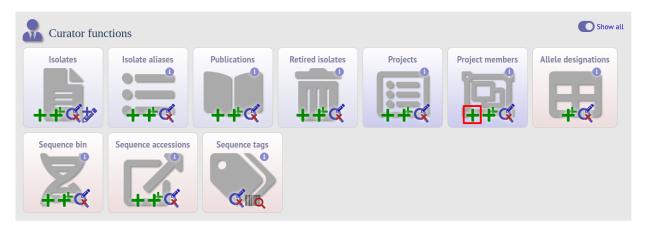
- private Setting this makes the project a private *user project*. You will be set as the project owner and will be the only user able to access it by default. You can add additional users or user groups who will be able to access and update the project data later.
- no\_quota If set, isolates added 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	<b>2</b>
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:! ○ true ● false 🕢	
private:! O true 🖲 false 🕢	
no quota:l 🖲 true 🔿 false 🕢	
restrict user:! ○ true ● false 🕢	
restrict usergroup:! 🔿 true 🖲 false 🕢	
curator:! Keith Jolley (keith) datestamp:! 2020-07-23	
<pre>full description: England, the Scottish Haemophilus, Legionella, Meningococcus and Pneumococcus Reference (SHLMPR) Laboratory, The Wellcome Trust Sanger Institute and the University of Oxford, funded by the Meningitis Research Foundation. Juse of the MRF Genome Library data <a href="/cgi-&lt;br&gt;bin/bigsdb/bigsdb.pl?db=pubmlst neisseria mrfgenomes">must be cited</a> in any publication or presentation making use of it.</pre>	
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	2 O
Record	Action
project id: 3) MRF Meningococcus Genome Library 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 +-#C	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 ( <b>include a field header line</b> ).	Action	
	RESET	

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		2
Instructions         Instage allows you to upload project member data as tab-delimited text or copied from a spreadsheet.         I ield header names must be included and fields can be in any order. Optional fields can be omitted if you wish. <b>Cemplates</b> Image: Im	Action RESET SUBMIT	

#### 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 spp. > Neisseria isolates > Isolate information		
Full information on isolate M17661 (id:193	377)	🔝 Help 🗹
Provenance/primary metadata		
isolate: M17661 strain designation: discrepancy: P1.5-1,10-8: F-ND: ST-11 (cc11) country: USA continent: North America region: MI year: 2008 disease: invasive (unspecified/other) species: Neisseria meningitidis serogroup: W	(71.17% cov),cssB fragmented       (16.16% cov),cssB fragmented       (17.2% cov),ctrA fragmented       (88.4% cov),ctrB fragmented       (98.45% cov),ctrE fragmented       date entered	<ul> <li>Keith Jolley, University of Oxford, UK</li> <li>Auto Tagger</li> <li>56 updates show details</li> <li>2012-06-27</li> <li>2020-02-10</li> </ul>
Secondary metadata  Vaccines		
Bexsero reactivity: insufficient data		
Versions More than one version of this isolate record exist. Newer versions: 40500 Publication (1)		
<ul> <li>Kislyuk AO, Katz LS, Agrawal S, Hagen MS, Conley AB, Jayaraman P, Nelakuditi \ IK (2010). A computational genomics pipeline for prokaryotic sequencing proje</li> </ul>		a 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:

²ubi∾	1LST															
me >	Organisn	ns > Neisse	ria spp. >	Ne	isseria isolate	s > Query or update	isolates									
Que	ery o	or upd	late i	SC	olates								E	Help	2](	
Enter	search o	riteria or lea	ave blank	to b	rowse all reco	ords. Modify form par	ameters to	filter or	enter a list of values.							
	Isolate p	provenance	fields						Display/sort option	IS						
ic			<b>~</b> =		✓ Ente	r value	+ (	0	Order by: id					<ul> <li>ascendi</li> </ul>	ng 🗸	
									Display: 25 🗸	records per page	(i)					
	Action_															
	RESET	SEA	RCH													
L	IL DE I															
83,41	.4 record	s returned (	1 - 25 disp	olaye	ed).											
—Та	ag scanr	ning	Projects						Private records	-Bookmark of	auerv					
	Scan	5	Select pro	ject			✓ Lin	k	Publish	2020-07-23:1		Ad	d bookmark			
$(\sim)$	Scan	12		ject 5	6 🕥	) (>>)	<ul><li>✓ Lin</li></ul>	k				Ad	d bookmark			
«	Scan			-	$\sim$	$\otimes$			Publish			Ad	d bookmark			
«	<ul> <li>(</li> </ul>	1 2		-	$\sim$	<b>&gt;</b>		k ate field	Publish				MLST		yping anti	igens
(«) Delete	Scan	1 2	34	-	$\sim$	allases			Publish		capsule					
(«) Delete	<ul> <li>(</li> </ul>	1 2 Sequence	3 4 New	5	6 📀		Isol	ate field	Publish	2020-07-23:1	capsule		MLST clonal	Finet		2 FetA_Vi
	<ul> <li>(</li> </ul>	1 2 Sequence bin	3 4 New version	5 id	6 🔊 isolate	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822;	lsol country	ate field year	Publish disease invasive (unspecified/other) meningitis and	2020-07-23:1 species Neisseria meningitidis Neisseria	capsule group		MLST clonal complex ST-1	Finet PorA_VR1	PorA_VR2	2 FetA_VI F1-5 🥖
×	<ul> <li>(</li> </ul>	1 2 Sequence bin	3 4 New version	5 id 1 2	6 > isolate A4/M1027	aliases B1; NIBSC_2803; Z1001	lsol country USA	ate field year 1937	Publish ds ds disease invasive (unspecified/other) meningitis and septicaemia invasive	2020-07-23:1 species Neisseria meningitidis Neisseria Neisseria	capsule group A	ST	MLST clonal complex	Finet PorA_VR1 5-2 Ø	PorA_VR2	2 FetA_VI F1-5 🥖
×	<ul> <li>(</li> </ul>	1 2 Sequence bin 1 2	3 4 New version + +	5 id 1 2	6 > isolate A4/M1027 120M	aliases B1; NIBSC 2803; Z1001 B35; NIBSC 2822; Z1035 B43; NIBSC 3076;	Isol country USA Pakistan	ate field year 1937 1967	Publish	2020-07-23:1 species Neisseria meningitidis Neisseria meningitidis Neisseria	capsule group A A	ST 1	MLST clonal complex ST-1 complex ST-4	Finet PorA_VR1 5-2 @ 5-2 @	PorA_VR2 10 2 10 2	2 FetA_VF F1-5 F5-1
×××	<ul> <li>(</li> </ul>	1 2 Sequence bin 1 1 2	3 4 New version + + +	5 id 1 2 3 4	6 ) isolate A4/M1027 120M M00242905 M1027	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035	Isol country USA Pakistan UK	ate field year 1937 1967 2000	Publish disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other)	2020-07-23:1 species Neisseria meningitidis Neisseria meningitidis	capsule group A A B	<b>ST</b> 1 1099	MLST clonal complex ST-1 complex	Finet PorA_VR1 5-2 🖋 5-2 🖋 19 🖋	PorA_VR2 10 🖋 10 🖋 15 🖋	2 FetA_VI F1-5 F5-1 +
× × × ×	<ul> <li>(</li> </ul>	1 2 Sequence bin ± ± ±	3 4 New version + + + + +	5 id 1 2 3 4 5	6 Solate A4/M1027 120M M00242905 M1027 M00240227	aliases B1; NIBSC 2803; Z1001 B35; NIBSC 2822; Z1035 B43; NIBSC 3076;	Isol country USA Pakistan UK USA UK	ate field year 1937 1967 2000 1937 2000	Publish disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other)	2020-07-23:1 species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group A A B A B B	<b>ST</b> 1 1099 4 1100	MLST clonal complex ST-1 complex ST-4 complex ST-32 complex	Finet PorA_VR1 5-2 @ 5-2 @ 19 @ + 7 @	PorA_VR2 10 @ 10 @ 15 @ + 16 @	2 FetA_VI F1-5 // F5-1 // + +
× × ×	<ul> <li>(</li> </ul>	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 4 5 6	6 isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207	aliases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043	Isol country USA Pakistan UK USA UK	ate field year 1937 2000 1937 2000 2000	Publish disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive	2020-07-23:1 specles Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Capsule group A A B A B B W	1 1099 4 1100 1101	MLST clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22 complex	Finet PorA_VR1 5-2 1 5-2 1 19 1 + 7 1 +	PorA_VR2 10 @ 10 @ 15 @ + 16 @ +	2 FetA_VF F1-5 0 F5-1 0 + + + +
× × × ×	<ul> <li>(</li> </ul>	1 2 Sequence bin ± ± ±	3 4 New version + + + + +	5 id 1 2 3 4 5	6 Solate A4/M1027 120M M00242905 M1027 M00240227	aliases B1; NIBSC 2803; Z1001 B35; NIBSC 2822; Z1035 B43; NIBSC 3076;	Isol country USA Pakistan UK USA UK	ate field year 1937 1967 2000 1937 2000	Publish disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive	2020-07-23:1 species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria	capsule group A A B A B B	<b>ST</b> 1 1099 4 1100	MLST clonal complex ST-1 complex ST-4 complex ST-32 complex ST-22	Finet PorA_VR1 5-2 @ 5-2 @ 19 @ + 7 @	PorA_VR2 10 @ 10 @ 15 @ + 16 @	2 FetA_VF F1-5 0 F5-1 0 + + + +
× × × × × ×	<ul> <li>(</li> </ul>	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 4 5 6 7	6 isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207	allases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043	Isol country USA Pakistan UK USA UK	ate field year 1937 2000 1937 2000 2000	Publish disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive	2020-07-23:1 specles Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria	Capsule group A A B A B 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 1 5-2 1 19 1 + 7 1 +	PorA_VR2 10 @ 10 @ 15 @ + 16 @ +	2 FetA_VF F1-5 0 F5-1 0 + + + +
× × × × × ×	<ul> <li>(</li> </ul>	1 2 Sequence 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 4 5 6 7	6 isolate A4/M1027 120M M00242905 M1027 M00240227 M00282207 7891	allases B1; NIBSC_2803; Z1001 B35; NIBSC_2822; Z1035 B43; NIBSC_3076; Z1043	Isol country USA Pakistan UK USA UK UK Finland	ate field year 1937 1967 2000 1937 2000 2000 1975	Publish disease invasive (unspecified/other) meningitis and septicaemia invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive (unspecified/other) invasive	2020-07-23:1 Species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group A A B A B W 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 @ 19 @ + 7 @ + 20 @	PorA_VRZ 10 @ 10 @ 15 @ + 16 @ + 9 @	PetA_VF           F1-5           F5-1           +           <

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	spp. > Neisseria isolates > Create new	isolate record version			
Create new iso	late record version				E Help C
designations will not. This fac	e a new version of the isolate record shov cilitates storage of different versions of ge y will be reset for the new record.				
— Enter new record id—	Options	Action			
id: 80312	Add new version to projects	CREATE			
Provenance/pr					
id:	5	genogroup:	A	sender:	Wendell Zollinger. Dept
id: isolate: aliases:	7	genogroup: genogroup notes:	A backbone: All essential capsule genes intact and present. Prediction code:		Wendell Zollinger, Dept Bacterial Diseases,Walter Reed Army Institute of Research, Washington DC, USA Auto Tanger
id: isolate: aliases: strain designation: country:	7 7891 B54; NIBSC_2760; Z1054 A: P1.20,9: F3-1: ST-5 (cc5) Finland		A backbone: All essential capsule genes intact and	curator:	Bacterial Diseases,Walter Reed Army Institute of Research, Washington DC, USA Auto Tagger
id: isolate: allases: strain designation: country: continent:	7 7891 B54; NIBSC_2760; Z1054 A: P1.20,9: F3-1: ST-5 (cc5) Finland Europe		A backbone: All essential capsule genes intact and present. Prediction code: https://github.com/ntopaz/chara cterize_neisseria_capsule.	curator:	Bacterial Diseases,Walter Reed Army Institute of Research, Washington DC, USA Auto Tagger 120 updates show details
id: isolate: allases: strain designation: country: continent: year:	7 7891 B54; NIBSC_2760; Z1054 A: P1.20,9: F3-1: ST-5 (cc5) Finland Europe 1975	genogroup notes:	A backbone: All essential capsule genes intact and present. Prediction code: https://github.com/ntopaz/chara cterize_neisseria_capsule. A	curator: update history: date entered:	Bacterial Diseases,Walter Reed Army Institute of Research, Washington DC, USA Auto Tagger 120 updates show details
id: isolate: aliases: strain designation: country: continent: year: disease:	7 7891 B54; NIBSC_2760; Z1054 A: P1.20,9: F3-1: ST-5 (cc5) Finland Europe 1975 invasive (unspecified/other)	genogroup notes: capsule group: MLEE designation: serotype:	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 2001-02-07
id: 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) CSF	genogroup notes: capsule group: MLEE designation:	A backbone: All essential capsule genes intact and present.Prediction code: https://github.com/ntopaz/chara cterize_neisseria_capsule. A Subgroup III 4,21 P1.9	curator: update history: date entered:	Bacterial Diseases,Walter Reed Army Institute of Research, Washington DC, USA Auto Tagger 120 updates show details 2001-02-07

# 6.25 Populating geographic coordinate lookup values

If a field has the geographic\_point\_lookup attribute set to 'yes' in the *config.xml file*, field values can be mapped to GPS coordinates to facilitate mapping.

If the user has curator privileges with the 'modify\_geopoints' permission set, or is an admin, a curator link to modify 'Geopoint field lookup' will be available in the curator interface. This link is normally hidden but can be shown by selecting the 'Show all' toggle.

To add a value, click the 'Add' link:



Select the 2 letter ISO country code and the linked field from the dropdown lists. Enter the town or city in the value field, and the GPS coordinates in LATITUDE, LONGITUDE format.

Public databases for mi and microbial genome						MY ACCOUNT
		HOME	ORGANISM	S SPECIES ID	ABOUT US	UPDATES
Home > Organisms > Neisseria spp. > Neisseria isolates >	<ul> <li>Add new geography point loo</li> </ul>	kup value				
Add new geography point le	ookup value				U	ser interface
Please fill in the fields below - required fields are marked	with an exclamation mark (!).					
Record		Action				
country code:! GB field:! town_or_city value:! Bristol location:! 51.45523, -2.59665 datestamp:! 2022-06-28 curator:! Keith Jolley (keith)	(1)	RESET	SUBMIT			
Contact	Cite us		т	follow	Suppo	orted by
				-0110 w	Suppo	nied by
Get in touch with us if you have any comments or suggestions concerning the website and the databases.	Please cite Jolley <i>et al.</i> 20 3:124 if you use data or an your publications.			9	Weikcom	e
Disclaimer & Privacy   Cookies   Terms & Conditions					Website b	y Manta Ray Media

Now, whenever that specified field value is used, a map will appear in the isolate information page showing the location. In addition, these values will be used in mapping in the Field Breakdown and Microreact plugins.

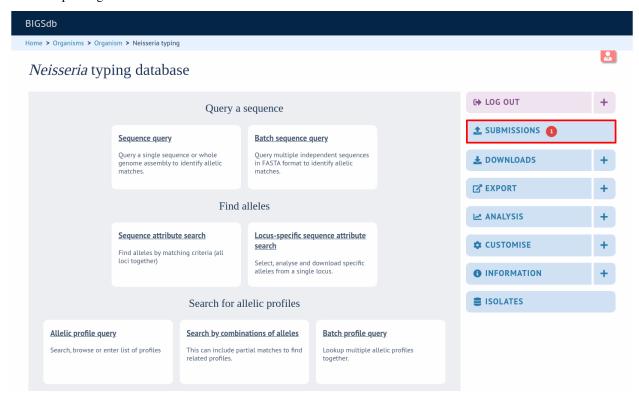
GPS coordinates can be mapped in bulk using the gp\_town\_lookups.pl script.

#### CHAPTER

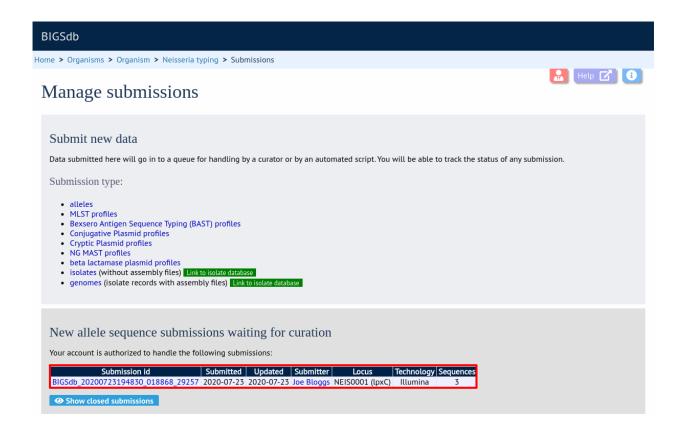
SEVEN

# 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.

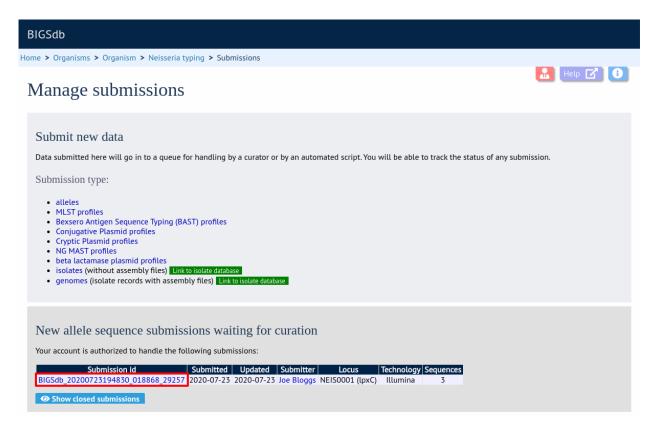


Any submissions for which you have sufficient privileges to curate will be shown.



# 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 CDS Status Query Assigned allele	Messages
UK322 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA $\checkmark$ pending $\checkmark$ $Q$ $\checkmark$ Curate UK323 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGATAA $\checkmark$ pending $\checkmark$ $Q$ $\checkmark$ Curate UK327 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA $\checkmark$ pending $\checkmark$ $Q$ $\checkmark$ Curate UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA $\checkmark$ pending $\checkmark$ $Q$ $\checkmark$ Curate BATCH CURATE BATCH QUERY	Message: Append Send now Archive Archive of submission and any supporting files: Download The

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	🔝 Hetp 🗗 🚺
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:     20-49x       assembly:     de novo       sequences:     Spades	
Sequences         Identifier   Length       Sequence       Complete CDS       Status       Query       Assigned allele         UK322       924       ATGCTGCAAAGAACTTTGGC        ACATCAGAACTTGGGATAA          ✓ pending          ✓ Q	Messages Message: Append Send now Archive
BATCH CURATE BATCH QUERY	Archive of submission and any supporting files:

## 7.1.1 Individual allele curation

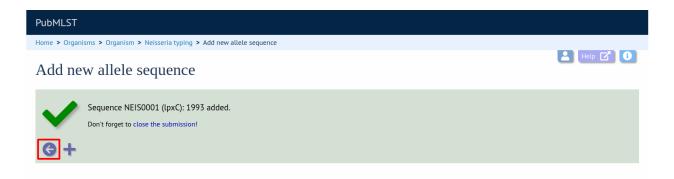
Individual sequences can be curated singly by clicking the 'Curate' links next to the sequence in the table. If you have supporting data attached to the submission, e.g. Sanger trace files then you may need to assess the submission based on the policy of the database.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	Help 🗹 🚯
Submission: BIGSdb_20200723194830_018868_29257	
type:       alleles         submitter:       Joe Bloggs, University of Oxford, UK         datestamp:       2020-07-25         status:       pending         locus:       NEIS0001 (lpxC)         sequences:       3         technology:       Illumina         read length:       100-199         coverage:       24-9x         assembly:       de novo         sequences:       Spades	
Sequences Identifier   Length   Sequence   Complete CDS   Status   Query   Assigner	Messages
UK322       924       ATGCTGCAAAGAACTTTGGC       ACATCAGAACTTGTGGATAA <ul> <li>pending</li> <li>Q</li> <li>Q</li></ul>	rate Message: Append Send now

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 > Orga	nism > Neisseria typing > Add new allele sequence	
Add new alle	le sequence	Help 🖍 🗿
Please fill in the fields belo	w - required fields are marked with an exclamation mark (!).	Action
locus:	NEIS0001 ~	
allele id:	1992 🖷	RESET SUBMIT
sequence:	ATCGATTTCGCCGGCAAATCCTACATCGACGAAATCGCGCGCG	
status:	unchecked 🗸	
	Bloggs, Joe (jbloggs)	
	Keith Jolley (keith)	
date entered:	2020-07-23 2020-07-23	
	O true O false New allele searches can be constrained to use just type alleles in comparisons	
comments		
Flags	alternative start codon a 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.



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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	Help 🗹 🚯
UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGATAA V pending Q Q 2Curate Mark all: Pending Rejected Update A	Messages tessage: Append Send now Archive crolive 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.

ne > Organisms > Organism > Neisseria typing > Submissions > Curate sub	mission	
urate submission		Help 🗹
Submission: BIGSdb 20200723194830 018868 292	57	
Summary		
type: alleles		
submitter: Joe Bloggs, University of Oxford, UK		
datestamp: 2020-07-23		
status: pending		
locus: NEIS0001 (lpxC)		
sequences: 3 FAS		
technology: Illumina		
read length: 100-199		
coverage: 20-49x		
assembly: de novo		
assembly software: Spades		
Sequences		Messages
Identifier Length Sequence	Complete CDS Status Query Assigned allele	3
UK322 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGA		
UK323 924 ATGCTGCAAAGAACTTTGGC ACATCAGAAATTGTGGA UK347 924 ATGCTGCAAAGAACTTTGGC ACATCAGAACTTGTGGA		Message: Append Send now
	Mark all: Pending Rejected Update	Archive
BATCH CURATE BATCH QUERY	opulate	Archive of submission and any supporting files:

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
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: NEIS0001 (lpxC)
sender: Blogs, Joe (blogs)
sequence JUK323 (FASTA):I ATGCTGCAAAGAATCCATCAGCGTTACCGGAGTCGGGCTGGATTCGGGC GAACGGCGCCGCGCC
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	Hetp 🕜 🕄
Sequences added. Don't forget to close the submission!	
<b>()</b> +	

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	
type       alleles         submitte:       Joe Bloggs, University of Oxford, UK         datestamp:       2020-07-23         status:       pending         locus:       NEISO01 (IpxC)         sequence:       3         sequence:       100-199         coverage:       20-49x         assembly softwar:       Spades	
Sequences       Messages         Identifier Length       Sequence       Complete CDS       Status       Query Assigned allele         UK322       924       ATGCTGCAAAGAACTTTGGC       ACATCAGAACTTGTGGATAA	
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	
Curate submission	A Help 🔽 🕚
Submission: BIGSdb_20200723194830_018868_29257  Summary  type: alleles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-23 status: pending outcome: accepted - data uploaded locus: NEIS0001 (lpxC) sequence: 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       ATGCTGCAAAGAACTTTGGC       Tejected v       Q       Curate         UK323       924       ATGCTGCAAAGAACTTTGGC       ACATCAGAAATTGTGGATAA       rejected v       Q       Curate         UK347       924       ATGCTGCAAAGAACTTTGGC       ACATCAGAACTTGTGGATAA       rejected v       Q       Curate         BATCH CURATE       BATCH QUERY       BATCH QUERY       Update       Update	Messages Message: Append Send now Archive Archive of submission and any supporting files: Download

## 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.

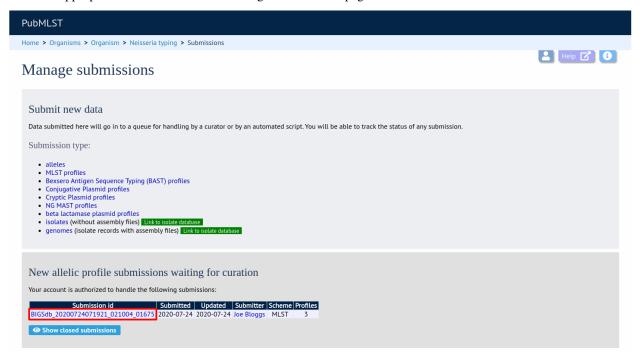
### 7.1.5 Closing the submission

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	🛓 Help 🕜 🚯
Submission: BIGSdb_20200723194830_018868_29257  Summary  type: alleles submitter: Joe Blogs, 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 software: Spades	
Sequences         Identifier Length       Sequence       Complete CDS       Status       Query Assigned allele         UK322       224       ATGCTGCAAAGAACTTTGGC        ACATCAGAACTTGTGGATAA         UK323       924       ATGCTGCAAAGAACTTTGGC        ACATCAGAAATTGTGGATAA         UK323       924       ATGCTGCAAAGAACTTTGGC        ACATCAGAAATTGTGGATAA         rejected ~       Q <td>N</td>	N
Message: Append Send now	

# 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 > C	Curate submission	
Curate submission		💄 Help 🕜 🚯
Submission: BIGSdb_20200724071921_02100	V4_01675 Profiles	
type: profiles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending	Identifier         abcZ         adk         arce         fumC         gdh         pdhC         pgm         Status         Query         Assigned ST           UK32         43         2         12         32         3         2         pending ♥         Q         Image: Curate           UK33         7         56         4         3         2         12         12         pending ♥         Q         Image: Curate           UK34         76         3         5         3         87         43         34         pending ♥         Q         Image: Curate           BATCH CURATE         Mark all:         Pending         Rejected         Update	
Messages Archive Archive of submis Download TAR Message: Append Send now	ssion and any supporting files:	

### 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	
Summary Profiles	
type: profiles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending BATCH CUR	Mark all: Pending Rejected Update
Messages Archive	
Archive of submission and any suppor Download The Message: Append Send now	files:

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

PubMLST		
Home > Organisms > Org	anism > Neisseria typing > Add new profile	
Add new M	LST profile	Help 🗹 🕚
Please fill in the fields b	low - required fields are marked with an exclamation mark (!).	
Record		
	15581	
abcZ: !		
adk: !		
aroE: !	12	
fumC: !	32	
gdh: !		
pdhC: !		
pgm: !		
	Bloggs, Joe (jbloggs) 🗸	
clonal_complex:	16-tal- 1-11 Restates	
date_entered: !	Keith Jolley (keith) 2020-07-24	
datestamp: !		
PubMed ids		
Action		
RESET	JBMIT	

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	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	💄 Help 🚺 🚯
Submission: BIGSdb_20200724071921_021004_01675Profiles	
type: profiles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending	
Messages Archive Archive of submission and any supporting files: Download Archive of submission and any supporting files:	

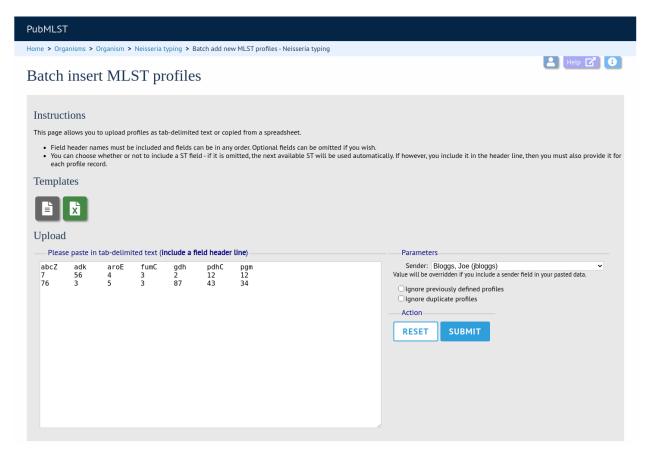
### 7.2.2 Batch profile curation

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

PubMLST	
Home > Organisms > Organism > Neisseria typing > Submissions > Curate submission	
Curate submission	L Help 🕜 🚯
Submission: BIGSdb_20200724071921_021004_01675	
Summary Profiles	
type: profiles submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending MARK all: Pending Rejected Update	
Messages Archive Archive of submission and any supporting files: Download TAR	
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 > O	Curate submission
Curate submission	Help 🔽 🕄
Submission: BIGSdb_20200724071921_02100	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         Ouerv/Assigned         ST           UK32         45         2         12         32         3         2         assigned         Q         15581           UK33         7         56         4         3         2         12         assigned         Q         15581           UK34         76         3         5         3         87         43         34         assigned         Q         15583           Messages         Archive         Archive         Archive         Archive         Archive
	Archive of submission and any supporting files:
	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 > 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 outcome: accepted - data uploaded	Identifier         abcZ         adk         aroE         fumC         gdh         pdhC         pgm         Status         Query         Assigned ST           UK32         43         2         12         32         3         2         rejected ∨         Q          ✓Curate           UK33         7         56         4         3         2         12         rejected ∨         Q          ✓Curate           UK34         76         3         5         3         87         43         34         rejected ∨         Q          ✓Curate           BATCH         CURATE         Mark all:         Pending         Rejected         Update
	Messages Archive
	Archive of submission and any supporting files:
	Download TAR
	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.

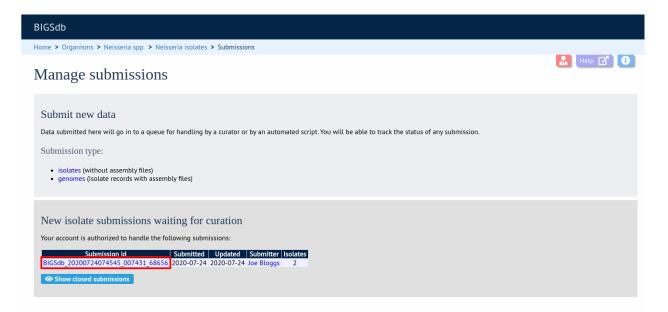
### 7.2.5 Closing the submission

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 > O	Curate 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         Messages	Identifier jabcZ jadk aroE fumC gdh pdhC pgm Status Query Assigned ST       UK32     43     2     12     32     3     2     rejected ×     Q       UK33     7     56     4     3     2     12     12     rejected ×     Q       UK34     76     3     5     3     87     43     34     rejected ×     Q       UK34     76     3     5     3     87     43     34     rejected ×     Q       Mark alt: Pending Rejected       Update	
Archive Action Archive of submission and any supporting files: CLOSE SUB Download	Message: Append Send now	

# 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 isola	tes > Submissions > Curate subr	nission		
Curate submission				Help 🕜 🚯
Submission: BIGSdb_20200724074	4545_007431_68656			
type: isolates submitter: Joe Bloggs, Universit datestamp: 2020-07-24 status: pending	ty of Oxford, UK			
Isolates				
isolate country year disease UK322 UK 2020 meningitis and septicaemia	a Neisseria meningitidis B	2 3 4 3 8	pdhC         pgm         PorA_VR1         PorA_VR2         FetA_VR         ST (N           4         6         5         2         F1-5         1	MLST) 11
UK325 UK 2020 septicaemia	Neisseria meningitidis B	2 3 4 3 8	4 6 5-1 2 F1-5 1 Record status: pending V Up	l1
BATCH CURATE			Record status, pending •	Juare
Messages	Archive			
	Archive of submission and any su	apporting files:		
	Download			
Message: Append Send now				

Click the 'Batch curate' button.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria iso	lates > Submissions > Curate submission	
Curate submission		Letp 🗹 🚯
Submission: BIGSdb_202007240 Summary submitter: Joe Bloggs, Univer datestamp: 2020-07-24 status: pending		
Isolates isolate country year disease UK322 UK 2020 meningitis and septicaem UK325 UK 2020 septicaemia	species serogroup abcZ adk aroE fumC gdh pdhC pgm PorA VR1 PorA VR ia Neisseria meningitidis B 2 3 4 3 8 4 6 5 2 Neisseria meningitidis B 2 3 4 3 8 4 6 5-1 2	R2 FetA_VR ST (MLST) F1-5 11 F1-5 11
BATCH CURATE	Record status:	pending V Update
Messages	Archive	
Message: Append Send now	Archive of submission and any supporting files:	

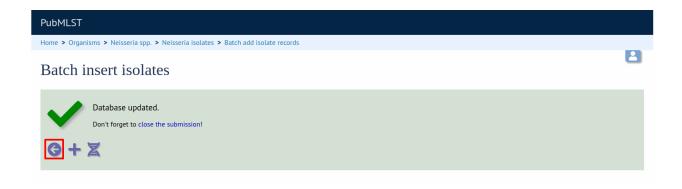
This will take you to the *batch isolate upload page* in the curators' interface.

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

PubMLST
Home > Organisms > Neisseria spp. > Neisseria isolates > Batch add isolate records
Batch insert isolates
Instructions         Inspage 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 () separated list.         • You can also upload allele fields along with the other isolate data - simply create a new column with the locus name (see the 'allowed_loci' tab in the Excel template for locus names). These will be used automatically.         • Ou can close whether or not to include an in number field - if it is omitted, the next available id will be used automatically. <b>Templates</b> Field header nof database fields for help with filling in the template. <b>Upload</b>
Please select the sender from the list below:
Bloggs, Joe (jbloggs) Value will be overridden if you include a sender field in your pasted data.
Paste in tab-delimited text (Include a field header line).     Action       isolate country year     disease species serogroup     abcZ     adk     aroE       fumC     gdh     pdhC     pgm     PorA_VR1     PorA_VR2     FetA_VR       UK322     UK     2020     meningitis and septicaemia     Neisseria meningitidis     B     Z       VK325     UK     2020     septicaemia     Neisseria meningitidis     B     Z     3       4     3     8     4     6     5 - 1     2     F1 - 5

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 Summary type: isolates submitter: Joe Bloggs, University of Oxford, UK datestamp: 2020-07-24 status: pending outcome: accepted - data uploaded	
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 septicaemia Neisseria meningitidis B 2 3 4 3 8 4 6 5 2 F1-5 11	2
UK325 UK 2020 septicaemia Neisseria meningitidis B 2 3 4 3 8 4 6 5-1 2 F1-5 11 Record status: pending Vupdate	
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:   Download	
Message: Append Send now	

CHAPTER

EIGHT

# **OFFLINE CURATION TOOLS**

## 8.1 Automated offline sequence tagging

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

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

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

sudo touch /var/log/bigsdb\_scripts.log
sudo chown bigsdb /var/log/bigsdb\_scripts.log

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

autotag.pl --database <database configuration>

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

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

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

A full list of options can be found by typing:

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

```
-d. --database NAME
   Database configuration name.
-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.
-1, --loci LIST
   Comma-separated list of loci to scan (ignored if -s used).
-L, --exclude_loci LIST
   Comma-separated list of loci to exclude
-m, --min_size SIZE
   Minimum size of seqbin (bp) - limit search to isolates with at least this
   much sequence.
-n, --new_only
   New (previously untagged) isolates only. Combine with --new_max_alleles
   if required.
--new_max_alleles ALLELES
   Set the maximum number of alleles that can be designated or sequences
   tagged before an isolate is not considered new when using the --new_only
   option.
-o, --order
   Order so that isolates last tagged the longest time ago get scanned first
    (ignored if -r used).
--only_already_tagged
```

```
Only check loci that already have a tag present (but no allele designation).
   This must be combined with the --already_tagged option or no loci will
   match. This option is used to perform a catch-up scan where a curator has
   previously tagged sequence regions prior to alleles being defined, without
   the need to scan all missing loci.
-p, --projects LIST
   Comma-separated list of project isolates to scan.
-P, --exclude_projects LIST
   Comma-separated list of projects whose isolates will be excluded.
-q, --quiet
   Only error messages displayed.
-r. --random
   Shuffle order of isolate ids to scan.
--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 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 scannew.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.
-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.
-q, --quiet
   Only error messages displayed.
```

(continued from previous page)

```
-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
   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 Calculating assembly stats

Basic assembly statistics are calculated automatically by the database engine as contigs are added to the sequence bin. These include the number of contigs, total length and the N50 value. Some calculations, such as %GC, number of Ns, and the number of gaps however, require offline analysis since these involve inspecting the nucleotide content of each contig. These can be calculated by running the update\_assembly\_stats.pl script. You can choose to run this against all databases on the system or against a specific database.

Only one copy of the script can run at a time, but it will stop gracefully if it detects another copy running, so it is recommended that the script is run regularly using a CRON job and the –quiet option. This ensures that records are updated shortly after they have been uploaded.

Once calculated, all assembly statistics can then be used in isolate queries.

A full list of options can be found by typing:

```
update_assembly_stats.pl --help
NAME
   update_assembly_stats.pl - Perform/update calculation of
    assembly GC, N and gap stats.
SYNOPSIS
   update_assembly_stats.pl [options]
OPTIONS
--database DATABASE CONFIG
   Database configuration name. If not included then all isolate databases
   defined on the system will be checked.
--exclude CONFIG NAMES
    Comma-separated list of config names to exclude.
--help
    This help page.
--quiet
    Only show errors.
--refresh_days DAYS
   Refresh records last analysed longer that the number of days set. By
   default, only records that have not been analysed will be checked.
```

## 8.5 Predicting species based on rMLST analysis

The *rMLST plugin* predicts species based on matches to rMLST alleles exclusively found in a particular species. It uses the PubMLST API to query either a genome sequence or rMLST allele designations to identify the species. When the analysis is run using the plugin, the results are also stored with the isolate record and can then be displayed within the isolate information page. This analysis can also be run offline using the update\_rmlst\_species.pl script.

Only one copy of the script can run at a time, but it will stop gracefully if it detects another copy running, so it is recommended that the script is run regularly using a CRON job and the –quiet option. This ensures that records are updated shortly after they have been uploaded.

A full list of options can be found by typing:

```
update_rmlst_species.pl --help
NAME
    update_rmlst_species.pl - Perform/update species id check
SYNOPSIS
    update_rmlst_species.pl [options]
OPTIONS
```

(continues on next page)

(continued from previous page)

```
--database DATABASE CONFIG
Database configuration name. If not included then all isolate databases
defined on the system will be checked.
--exclude CONFIG NAMES
Comma-separated list of config names to exclude.
--help
This help page.
--last_run_days DAYS
Only run for a particular isolate when the analysis was last performed
at least the specified number of days ago.
--quiet
Only show errors.
--refresh_days DAYS
Refresh records last analysed longer that the number of days set. By
default, only records that have not been analysed will be checked.
```

## 8.6 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

## 8.7 Uploading contigs from the command line

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

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

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

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

A full list of options can be found by typing:

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

# 8.8 Populating geographic point lookup values

If a field has the geographic\_point\_lookup attribute set to 'yes' in the *config.xml file*, field values can be mapped to GPS coordinates to facilitate mapping.

These lookup values can be populated using the gp\_town\_lookups.pl script found in the scripts/maintenance directory. This script uses the Geonames dataset that contains GPS coordinates for towns and cities with populations of at least 1000 people. The dataset is included in the datasets/Geonames directory.

A full list of options can be found by typing:

```
gp_town_lookups.pl --help
NAME
   geography_point_lookups.pl - Populate geography_point_lookup table
   to set city/town GPS coordinates for mapping.
SYNOPSIS
   geography_point_lookups.pl --database NAME --field FIELD --geodataset DIR
   Run this to populate any unassigned values in the geography_point_lookup
    table.
OPTIONS
--database NAME
   Database configuration name.
--feature_code CODE
   Geonames feature code. See http://www.geonames.org/export/codes.html.
   Default is 'P' (towns/cities).
--field FIELD
   Name of field. This should have the geography_point_lookup attribute set to
    'yes' in config.xml.
--geodataset DIR
   Directory containing the Geonames dataset.
--help
   This help page.
--min_population POPULATION
   Set the minimum population for town to assign. Note that all entries in the
   Geonames database has population, so setting this attribute may result in
    some values not being assigned, but can ensure that only high-confidence
   values are used.
--quiet
   Only show error messages.
--tmp_dir DIR
   Location for temporary files. Defaults to /var/tmp/.
```

### CHAPTER

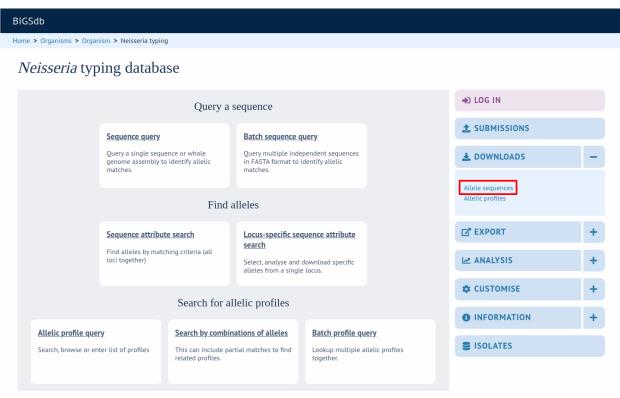
### NINE

# **DEFINITION DOWNLOADS**

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

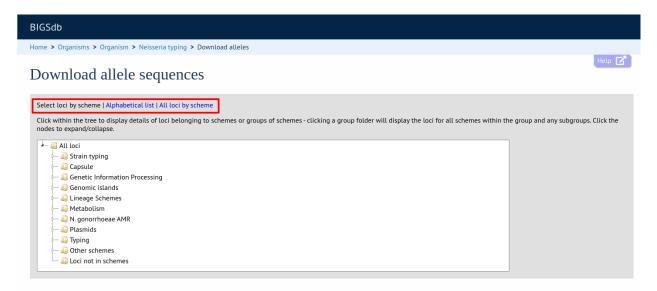
# 9.1 Allele sequence definitions

Click the 'Allele sequences' link in the 'Downloads' section.



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

### 9.1.1 Scheme tree



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

BIGSdb							
Home > Or	ganisms	<ul> <li>Organis</li> </ul>	m > Neisseria typir	ng > Downlo	ad alleles		
Dowr	nload	l alle	le sequen	ices			Help 🗹
Select loo	i bv schen	ne   Alpha	betical list   All loci	by scheme			
	in the tree	e to displa			nemes or groups	f schemes - clicking a group folder will display the I	loci for all schemes within the group and any subgroups. Click the
	Strain ty Capsule Genetic Genomic Lineage Metabol N. gonor Plasmid: Typing Jup Finet	Information islands Schemes ism rhoeae Al s yping ant					
			es length (setting)	Min Ioneth	Max Jonath Full	ame/product Curator(s) Last updated	
abcZ		DNA 107		433	434	O. Harrison 2020-06-23	
adk		DNA 84		464	465	O. Harrison 2020-07-14	
aroE	۵ 🛓	DNA 111	8 Fixed: 490 bp	489	493	O. Harrison 2020-06-12	
fumC		DNA 118		463	467	O. Harrison 2020-06-12	
gdh		DNA 115		501	513	O. Harrison 2020-07-03	
pdhC		DNA 109		480	503	O. Harrison 2020-06-04	
pgm	۵ 🛓	DNA 117	2 Fixed: 450 bp	447	489	O. Harrison 2020-06-12	
Ľ	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	-	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
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGG
TTGGGCGAAATCAAAATCCCCGAAAACGAGCCGGGTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGTATGGCGGGCGCGCGGCGCAATTTCGAGCTGAACGTCTATATGCCCGTCATC
GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCGTCGGCATTGAACCCGTACCGGAAAAAATCGACTATTTCCTGCACCATTCC
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAC
>fumC_2
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGG
TTGGGCGAAATCAAAATCCCCGAAAACGAGCCGGGTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGCATGGCGGGCGCGCGCGGGCAATTTCGAGCTGAACGTCTATATGCCCGTTATC
GCCTACAACCTCTTGCAATCCATCCGCCTCTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCATCGGCATCGAACCCGTACCGGAAAAAATCGACTATTTCCTGCACCATTCC
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAC
>fumC_3
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGG
TTGGGCGAAATCAAAATCCCCGAAAAACGAGCCGGGTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGCATGGCGGGCGCGCGCGGGCAATTTCGAGCTGAACGTCTATATGCCCGTTATC
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									
ne > Organisms	> Organisı	m > N	eisseria †	typing > Download alleles					
ownload	d alle	le s	equ	ences					Help 🔽
elect loci by sche	malAlpha	hetical	list   All	loci by scheme					
	me   Atphat	Deticat	ust   Au	toer by scheme					
-									
Locus Downl	oad Type A		Leng	th (setting) Min length Ma	v length Eu	ull name/produc	t Last updated		
	Joau Type /	Alleles	Leng	ui (setuing) (mini tengun ma					
23S rRNA 🛛 📥	DNA	1101	Variable	No limits set 2884	2966		2019-08-23		
23S_rRNA 📥	DNA	1101	Variable	e: No limits set 2884	2966				
235_rRNA 🛃	DNA	1101	Variable	: No limits set 2884	2966				
	DNA			: No limits set 2884 Length (setting)	Min	Max		Curator(s)	Last
Ą							2019-08-23	Curator(s)	Last updated 2020-05-2
Locus	Download	Туре	Alleles	Length (setting)	Min length	Max length	2019-08-23 Full name/product	Curator(s)	updated 2020-05-2
Locus ArsR [NEIS1769]	Download	Type DNA DNA	Alleles 195	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 2020-05-2 2020-03-2
Locus ArsR [NEIS1769] AsnC [NEIS1566]	Download	Type DNA DNA DNA	Alleles 195 394	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		updated 2020-05-2 2020-03-2 2020-06-2
Locus ArsR [NEIS1769] AsnC [NEIS1566] abcZ	Download	Type DNA DNA DNA DNA	Alleles 195 394 1077	Length (setting) Variable: No limits set Variable: No limits set Fixed: 455 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 2020-05-2 2020-03-2 2020-06-2 2020-07-1
Locus ArsR (NEIS1769) AsnC (NEIS1566) abcZ abcZ (NEIS1015)	Download	Type DNA DNA DNA DNA	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	Full name/product Full name/product ArsR family transcriptional regulator transcription regulator AsnC ABC transporter ATP-binding protein		updated 2020-05-2 2020-03-2 2020-06-2 2020-07-1 2020-07-1
Locus ArsR (NEIS1769) AsnC (NEIS1566) abcZ abcZ (NEIS1015) aceF (NEIS1079) ackA2	Download	Type DNA DNA DNA DNA DNA	Alleles 195 394 1077 1263 1550	Length (setting) Variable: No limits set Variable: No limits set Fixed: 433 bp Variable: No limits set Variable: (1563 min; 1641 max)	Min length 270 459 433 1890 1533	Max length 291 471 434 1944 1641	Full name/product Full name/product ArsR family transcriptional regulator transcription regulator AsnC ABC transporter ATP-binding protein dihydrolipoamide acetyltransferase (EC 2.3.1.12)		updated           2020-05-2           2020-03-2           2020-06-2           2020-07-1           2020-07-1           2020-06-1
A ArsR (NEIS1769) asnc (NEIS1566) abcZ (NEIS1015) aceF (NEIS1279) ackA2 [NEIS1727]	Download	Type DNA DNA DNA DNA DNA DNA	Alleles 195 394 1077 1263 1550 762	Length (setting) Variable: No limits set Variable: No limits set Fixed: 433 bp Variable: No limits set Variable: (1563 min; 1641 max) Variable: No limits set	Min length 270 459 433 1890 1533 1197	Max length 291 471 434 1944 1641 1212	Full name/product         ArsR family transcriptional regulator transcription regulator AsnC         ABC transporter ATP-binding protein dihydrolipoamide acetytransferase (EC 2.3.1.12)         acetate kinase		updated
A Locus ArsR (NEIS1769) abc2 (NEIS1566) abc2 (NEIS1015) aceF (NEIS1279) accA2 (NEIS1727) accA (NEIS1727)	Download	Type DNA DNA DNA DNA DNA DNA	Alleles 195 394 1077 1263 1550 762 1211	Length (setting) Variable: No limits set Variable: No limits set Fixed: 433 bp Variable: No limits set Variable: 1563 min; 1641 max) 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	Full name/product 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           2020-05-2           2020-03-2           2020-06-2           2020-07-1           2020-07-1           2020-06-1           2020-06-1           2020-05-2

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.

IGSdb	)												
ome > O	Drganism	1s > 0	rganism	> Neis	sseria typin	g > Downloa	ad alleles						
													Help 🔽
	nla	d a	11.1		auon								
JOW	IIIOa	iu a	illei	e se	equen	ces							
Select Io	nci by sci	homo I	Alphabe	tical lie	st   All loci	hyscheme							
Select to	oci by sci	lienie j	мирнарс	incar in:		by schenie							
MLS	Г												
Locus D	Downloa	id Type	Alleles	Lengt	h (setting)	Min length N	1ax length	Full name/p	oroduct Curator(s) L	ast updated			
abcZ	*		1077		d: 433 bp	433	434		O. Harrison				
adk			841		d: 465 bp	464	465		O. Harrison				
aroE			1118		d: 490 bp	489	493		O. Harrison				
fumC	*		1184		d: 465 bp	463	467		O. Harrison				
gdh	4		1156	Fixed	d: 501 bp	501	513		O. Harrison				
pdhC	*		1096	Fixed	d: 480 bp	480	503		O. Harrison				
pgm	*	DNA	1172	Fixed	d: 450 bp	447	489		O. Harrison	2020-06-12			
Finety					L				E II and the day	Company (1)			
Locus PorA VR		nload	Type peptide	Alleles 360		(setting) No limits set	Min length 9	Max length 31	Full name/produc		Last updated 2020-06-30		
PorA VR			peptide	988		No limits set	5	41	PorA variable region		2020-06-30		
FetA V			peptide			No limits set	20	69	FOR Variable region		2020-06-08		
TCUN_V	× •	in I	reptide	055	variable.	to timits set	20	09		1. I cavers	2020-00-00		
rplF s	pecie	S											
	Downloa	d Type	Alleles	lenat	h (settina)	Min length N	lax length	Full	l name/product	Curator(s)	Last updated		
rplF			257		d: 413 bp	413			nal protein L6 fragme				
1940	-	510	207	. 1/(0)					int protein Lo nogine		2020 35 10		

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.

	Download						Full name/product		
abcZ	-	DNA	1077	Fixed: 433 bp	433	434			2020-06-23
adk	<b>.</b>	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	*	DNA	1156	Fixed: 501 bp	501	513		O. Harrison	2020-07-03
pdhC	<u>.</u>	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 > Orga	nism > Neisseria typin	g					
<i>Neisseria</i> typ	oing databa	se					
		DILOG IN					
	Sequence query		Batch sequence of	<u>uery</u>			
	Query a single seque genome assembly to matches.		Query multiple inde in FASTA format to matches.			Ł DOWNLOADS	-
		Find a	lleles			Allele sequences Allelic profiles	
	Sequence attribut		Locus-specific sequence attribute search				+
	Find alleles by matc loci together)	hing criteria (all	Select, analyse and download specific alleles from a single locus.				+
		Search for all	lelic profiles				+
							+
Allelic profile que Search, browse or en		Search by combination of the search by combination of the search by combination of the search by combined profiles.		Batch profile que Lookup multiple a together.		SISOLATES	
				Ū.			

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 update
MLST	*	15,577		Odile Harrison, Keith Jolley	2020- 07-14
rplF species	*	211		James Bray, Ellie Watkins, Julia Bennett, Kanny Diallo	2019- 02-06
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- 07-13
Conjugative Plasmid	*	242		Ana Cehovin	2018- 04-25
Cryptic Plasmid	*	96		Ana Cehovin	2018- 05-15
GC OMV peptide typing	*	2,283	OMV peptide typing based on most abundant components of gonococcal OMVs	Anastasia Unitt, Odile Harrison	2020- 03-31
N. gonorrhoeae cgMLST v1.0	*	8,453			2020- 07-14
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- 07-14
NG MAST	*	19,141	N. gonorrhoeae typing scheme using porB and tbpB fragments. In development, please DO NOT USE		2020- 06-25
NG STAR	*	1,933	Neisseria gonorrhoeae Antimicrobial Sequence Type In development, please DO NOT USE	Odile Harrison	2020- 07-14
OMV peptide typing	*	4,046	OMV peptide typing based on the components of the MeNZB vaccine	Hannah Chan, Charlene Rodrigues	2019- 05-15

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

ST	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	clonal_complex
1	1	3	1	1	1	1	3	ST-1 complex/subgroup I/II
2	1	3	4	7	1	1	3	ST-1 complex/subgroup I/II
3	1	3	1	1	1	23	13	ST-1 complex/subgroup I/II
4	1	3	3	1	4	2	3	ST-4 complex/subgroup IV
5	1	1	2	1	3	2	3	ST-5 complex/subgroup III
6	1	1	2	1	3	2	11	ST-5 complex/subgroup III
7	1	1	2	1	3	2	19	ST-5 complex/subgroup III
8	2	3	7	2	8	5	2	ST-8 complex/Cluster A4
9	2	3	8	10	8	5	2	ST-8 complex/Cluster A4
10	2	3	4	2	8	15	2	ST-8 complex/Cluster A4
11	2	3	4	3	8	4	6	ST-11 complex/ET-37 complex
12	4	3	2	16	8	11	20	
13	4	10	15	7	8	11	1	ST-269 complex
14	4	1	15	7	8	11	1	ST-269 complex

### CHAPTER

TEN

## **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 ii	nformatio	on on is	olate M10 2	240474 (id:1	18968)		Help
				× ×	,		
📃 Pr	ojects						
This isolat	e is a member of th	e following pro	ject:				
The MRE	Meningococcus Ge	nome Library is	a collaboration between		coccus Genome Library ne Scottish Haemophilus, Legionella,	Meningococcus and Pneumog	
Reference	e (SHLMPR) Laborat	ory, The Wellco	me Trust Sanger Institute		ford, funded by the Meningitis Resea		Meningitis 2 Research Foundation
🌔 Pr	ovenance/pr	imary met	adata				
		18968			invasive (unspecified/other)	update history:	223 updates show details
5	isolate: strain designation:	M10 240474 B: P1.19-1,15-	11: F3-9: ST-269	species: serogroup:	Neisseria meningitidis B	date entered: datestamp:	2012-02-15 2020-02-28
	-	(cc269)		genogroup:	В	a a contraction of the second s	
	country: continent:	UK [England] Europe		capsule group: ENA run accession:	B ERR086224 www.ebi.ac.uk		
	-	South East			Dorothea Hill, University of Oxford,		
epie	year: demiological year:		011	curator:	UK Auto Tagger		
Se	econdary met	adata					
🔎 V	/accines						
1	Bexsero reactivity:	cross-reactive	8	Trumenba reactivity:	cross-reactive	Trumenba notes:	fHbp_peptide: 15 is cross-reactive to vaccine variant - data derived
	Bexsero notes:	to vaccine var from MATS as (PMID:234147 PMID:2668699 PMID:2735562	09, PMID:23588089, 98, PMID:26950303, 28, PMID:28366725, 18, PMID:30592763,		Inotes		from MEASURE assays (PMID:22553195), and SBA assays (PMID:2259484, PMID:22718089, PMID:225352429, PMID:26407272, PMID:26707218, PMID:26803528, PMID:2635794, PMID:27745812, PMID:27846061, PMID:28196734, PMID:28566335, PMID:29236639)
	ublications (4		,				
	Brehony C, Rodri	gues CM, Borro	w R, Smith A, Cunney R, N nisation. <i>Vaccine</i> <b>34</b> :4690		16). Distribution of Bexsero® Antigen	sequence Types (BASTs) in ir	wasive meningococcal disease
					olley KA, Bratcher HB, Parkhill J, Tang ort study. <i>Lancet Infect Dis</i> <b>15:</b> 1420-8		015). Genomic epidemiology of age-
				Wheeler J, Whiting G, Feav Open Res <b>3:</b> 151 3505 is	rers IM, Maiden MCI (2018). Typing co olates	mplex meningococcal vaccine	es to understand diversity and
				n JC, Moxon ER, Maiden M rg Infect Dis <b>24:</b> 673-682		MenB Vaccine Antigenic Varia	ants among Disease-Causing Neisseria
🗶 Se	equence bin						
	max length:	2,195,045 bp 109,859 bp	N50 contig number: N50 length (L50): N90 contig number:	34,308 N95 74	ntig number: 97 length (L95): 3,513 loci tagged: 2,179		
SI	mean length: how sequence bin	7,982 bp	N90 length (L90):	0,100			
	chemes and l	oci					
	Capsule Genetic Informatic Genomic islands Lineage Schemes Metabolism Pilin Typing Other schemes		Navigate and select sch	emes within tree to displa	ay allele designations		
	Loci not in scheme	25					

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 Pneumococc 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.	CCUS 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 tanged. Senbin id: 4913: 18651-19082 (sequence bin scan)

## **10.1.3 Publications**

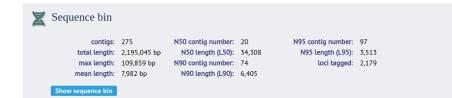
#### Publications (4)

- 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					Μ	LST			
🛺 Genetic Information Processing	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST	clonal complex
🖙 🚚 Genomic islands	<b>4</b> 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									
🛺 Finetyping antigens									
🚚 16S									
🚚 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	Help 🗹
length: status: date entered: datestamp: sender:	abcZ 2 TITGATACCG TIGCCGAAGG TITGGGTGAA ATTCGCGATT TATTGCGCCG TTACCACCGC GTCGGCCATG AGTTGGAAAA CGGTTCGGGT GAGGCT TGAAAGAACT CAACGAATTA CAACTTGAAA TCGCAAGCGAA GGACGGGCTG CGGCAGTCAA GCAGACTTG GGGGAACTCG GTTG AAACGAAAAA ATTGGCAACCT TTCCGGCGG TCAGAAAAAG CGTGTGCGCT TGGCGCAGGCAC TGGGTGGCAG AAGCCGACG CGGCAGCCG ACCAACCAATA ATTGGCACGCG TCAGAAAAAG CGTGTGCCT TGGCGCAGGACCTGG TTGTGATTAC CCACGACCGC CGTTT ACAATATCGC CACGCGGATT GTCGAACTCG ATC 433 Sanger trace checked	CGGA
Profiles contai		
	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 databas one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represe sample. <u>11940 isolates</u>	

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	Help 🗹
Contig position sequence bln ld: 4956 contig length: 22993 start: 13864 end: 14607 length: 744	Options Flanking sequence length: 100 v Reload
orientation: forward complete: yes method: Illumina	
Sequence CACTGTACGC GGCGCACGGC TTCAGCATTG CGGGCAGGCG GAAAAACTAT TACCGTACAG CCGACGGTAA AACCGAAGAT GCCGTCTTAA TGGAGAA CCACCTGCAC GAAGCCTTGG GTTTGGGCCC GATGTGGTG AAGCGGGGAG CCGCGCCTG GCCGTCGCGA ACATTGCCGG AAAGCCGCGA ACAATTGCCGG CGGCGGCGT TCAGACGGCA CCGCACCGCT TCCGGCGCATA CCAGACGTCG AAACTGCGCG TACATACGCG GTCCGGCGTT TCAGACGGCA TCCGCCCCGT TCCGCCGCGT TCGGGCATAA CCAGACTGC GCTGTGTCAGC CTGGTCGGC CGACCCGC GTCCGGCGGTT TCAGACGGCA ATATACTCAA AGCCGTAGG CTCGATGCCG CCTATGTCCG CGCAGCCGTCG GACGCCG CGCAGTGCC CTGCTCGACAA ATATACTCAA AGCCGTAGG CTCGATGCCG CCTATGTCCG CCAAACCTG CGCAGATGCG CTGGCAACAAA CCGCCCCGCGA ATATACTCA AGCCGTGGCGCGCC GGCCGCCCC TTCTTCATCA ATGCCCCCGCGA CTCGAGGCT TGCGCCGCCGCCCGCCGCGCCG	CGE CEGEAAAAGC AAACCGTCCT GGG CCTGAAACGG AAACGCCTCT TAC GGGCAACTGT TCCACGGAAA CCG ATGCCGTCT AAGCGGCAAT GGG ATGATTGAAA CTTTGTGCGC GTC TTGGCGCAAG GCGGCGGCAG
H C T R R T A S A L R A G G K T I T V O 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 C 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 CACTGRACGCGCCCACCGCTTCACCATTGCGGCAGGCGGAGAAACTATTACCGTACAGCCGACGAGAAACCGAACATGCCGCCCTTATGGAGAAAT 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 S V R P N S I P V R H G S K P K F3         201 AAAGCCCGACAAATCCGCCCCGAAAAGCAAACCGTCCTCAGCATCCGGCCGCGACAGCATCCGGTCAGGACAGGCTCAAAACCATGAA 300	

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

Sequence bin	for M10 24047	<i>'</i> 4			Help 🔽 🧯
Contig summary s Contigs Total length Minimum length Maximum length N50 contig number N50 length (L50) N90 contig number N90 length (L90) N95 contig number N95 length (L95)	130       131       132       133       134       135       136       137       138       139       130       131       131       132       133       134       135       136       137       138       139       130       131       131       132       133       134       130       145       130       145       130       145 </th <th>Contig size distribution</th> <th>0000 Contig</th> <th>275</th> <th></th>	Contig size distribution	0000 Contig	275	
FASTA EMBL GBK	CFF3	٩		(Cursor Guide) Track L	abels Save SVG *
	4811 4818 4836	4869	4917 4921 4947 4953 4957	4959 4989 5000 5	005
promoter region pro_NEIS0488			pro_NEIS1635 pro_NEIS0350		
serC aspA carB	pip ppk pilA dhpt	S 'porB adk gyrA	fHbp_DNAfrag_Pasteur	IrplF talA gpm	рукА
NEISp0612 NEISp219	98 NEISp0715 NEISp1364	NEISp0408 NEISp1130 NEISp0210 NEISp1783 NEISp1063		J96NEISp0101 NEISp1503 NEISp0531	NEISp1687
partial protein	PorA_VR3		Hbp_peptide Hbp_Peptide Hbp_PEPTIDEIrag_Pastet	FetA_VR	\$
intergenic region	igr_up_NEIS2109	) igr_u	_NEIS0055 igr_up_NEIS0351		*
RNA25a IRNA	A-leu tRNA-arg tRNA-an I A-pro	g NG_ponA IRNA-leu IRI	igr_upNEIS0350	p tRNA-trp tRNA-phe tRNA-gly tRNA	A-gin hmbR
					÷
Sequence Sequencing met 4869 Illumina	thod Original designa NODE_90_length_109787_r		Locus         Star           NEIS1151         488           NEIS1150         114           NEIS1149         3155           NEIS1148         3899           NEIS1147         452           NEIS1146         5399           NEIS1145         614           NEIS1144         6893	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	

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

### ELEVEN

## **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 DI LOG IN Find alleles Search for allelic profiles Query a sequence By specific criteria By specific criteria Single sequence Search, browse or enter list of profiles Ouery a single sequence or whole Find alleles by matching criteria (all LOWNLOADS genome assembly to identify allelic loci together) atches. EXPORT Batch sequences By locus By allelic profile MALYSIS + Select, analyse and download specific This can include partial matches to find Ouery multiple independent sequences in FASTA format to identify allelic alleles from a single locus. related profiles CUSTOMISE + matches. **1** INFORMATION + <u>In a batch</u> Look up multiple allelic profiles together. SISOLATES

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 fire nearest partial matches will be identified if an exact match is not found. You can query using eithe Please select locus/scheme Order results by MLST I Iocus I Construction of the constr	

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. ()
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
1 exact match found. Translate query
Locus Allele Length Contig Start position End position Flags Comments arot 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.
Initial of the second secon
Translate query Closest match: aroE: 64
Show alignment
Differences 2 differences found. $$ $5^{5}A \rightarrow ^{51}T$
$^{134}A \rightarrow ^{230}C$ 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	3
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.	
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.

ome > Orga	inisms >	Organism	> Neisseria t	yping > Sequ	uence query		
eque	nce c	luery					Help 🗹 🚺
							an exact match against the chosen (or all) loci - they do not need to be trimmed. The A or peptide sequences. ()
Please	select lo	ocus/scher	ne		Order results by		
MLST				~	locus 🗸		
Enter	query sec	quence (sir	ngle or mult	tiple contigs	s up to whole genom	e in size)	Alternatively upload FASTA file
							Select FASTA file:
							Choose file No file chosen
							or enter Genbank accession — Action —
							RESET SUBMIT
/ exact mat							
7 exact mat	le Length		Start position 5637	End position	n Flags Comments		
Locus Alle abcZ 2 adk 3	le Length 433 465	<b>Contig S</b> 187414 187432	5637 8935	6069 9399	n Flags Comments		
Locus Alle abcZ 2	le Length 433 465 490	<b>Contig</b> S 187414	5637	6069	n Flags Comments		
Locus Alle abcZ 2 adk 3 aroE 4 fumC 3 gdh 8	le Length 433 465 490 465 501	Contig S 187414 187432 187449 187396 187466	5637 8935 4826 4871 7333	6069 9399 5315 5335 7833	n Flags Comments		
Locus Alle abcZ 2 adk 3 aroE 4 fumC 3 gdh 8 pdhC 4	Length 433 465 490 465 501 480	<b>Contig S</b> 187414 187432 187449 187396	5637 8935 4826 4871	6069 9399 5315 5335	n Flags Comments		
Locus Alle abcZ 2 adk 3 aroE 4 fumC 3 gdh 8 pdhC 4	Length 433 465 490 465 501 480	Contig S 187414 187432 187449 187396 187466 187542	5637 8935 4826 4871 7333 56831	6069 9399 5315 5335 7833 57310	n Flags Comments		
Locus Alle abcZ 2 adk 3 aroE 4 fumC 3 gdh 8 pdhC 4 pgm 6	Le Length 433 465 501 480 450 450 450	Contig S 187414 187432 187449 187396 187542 187542 187395	5637 8935 4826 4871 7333 56831	6069 9399 5315 5335 7833 57310	n Flags Comments		

# 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.

BIGSd	-				
	rganisms > Organism > Neisseria typing				
Νε	<i>eisseria</i> typing databas	se			
	Query a sequence	Find alleles	Search for allelic profiles	DILOG 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	+
	induites.			EXPORT	+
	Batch sequences Query multiple independent sequences	By locus Select, analyse and download specific alleles from a single locus.	By allelic profile This can include partial matches to find	MALYSIS	+
	in FASTA format to identify allelic matches.		related profiles.		+
	J		In a batch	<b>1</b> INFORMATION	+
			Look up multiple allelic profiles together.	SISOLATES	

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.  Please select locus/scheme Order results by MLST I locus V	
Enter query sequences (FASTA format) Alternatively upload FASTA file	
>sample_1 GAAGCGAÄAAAATCATTGACGAAGGCGGCTTGGTGCGCGACGACATCATTATCGGCATGGTCAAAGAAC GCATCGCGCAAGACGACTGCAAAAACGGTTTCCTGTTCGACGGCGTTTCCCGCGCACAATTGGCACAAGCCGA AGCGATGGTTGAAGCAGGCGTGGATTTGGATGCAGTGCGTGAAATCGACGTGCCTGACAGCGGGATTGTC GACCGTATGAAGCCGCCGCCGCGTGCATTTGGCTTCCGGCCGTACTTACCACGGTGACGACGACGACGACGACGACGACGACGACGACGACGACG	

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

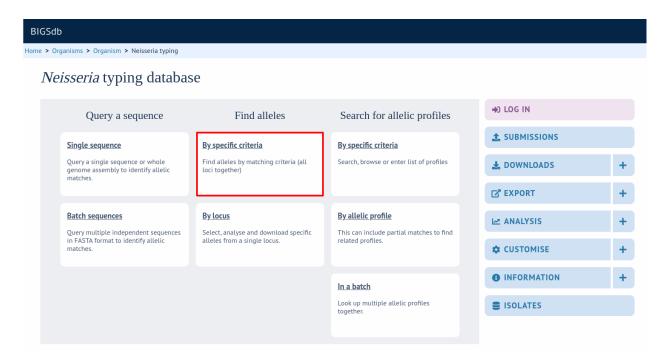
BIGSdb	
Home > Organisms > Organism > Neisseria typing > Batch sequence query	
Batch sequence query	8
Please paste in your sequences to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences. 🕢	
Please select locus/scheme Order results by	
MLST V locus V	
Enter query sequences (FASTA format) Alternatively upload FASTA file	
<pre>&gt;sample_1 GAAGCGAAAAAATCATTGACGAAGGCGGCTTGGTGCGCGACGACATCATTATCGGGCATGGTCAAAGAAC GCATCGCGCAAGACGACTGCAAAAACGGTTTCCTGTTCGACGGCTTGCCAGCGCGCACATTGGCACAGCCGA AGCCGTTGAAGCCAGCGGTGGATTGGAT</pre>	
Contig       Match       Locus       Allele       Differences         sample_1       exact       adk       3         sample_2       partial       adk       21       1 difference found. <sup>237</sup> T - <sup>237</sup> A         sample_3       partial       adk       27       2 differences found. <sup>230</sup> G - <sup>230</sup> A; <sup>425</sup> A - <sup>425</sup> C	

# **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.



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 s export plugins.	equence analysis or
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 unle search to a locus that uses integer allele ids using the drop-down list.	ss you filter your
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 d settings.	lisplay and query
Search criteria	
locus     •     =     •     abcZ     •       locus     •     ascending •       Display:     25 •     records per page (j)	

Click 'Search'.

BIGSdb				
Home > Organisms	> Organism >	Neisseria typing > Sequend	e attribute search	
Sequence	attribut	te search		Hetp 🗗 🚺
Some loci have ad export plugins.	lditional fields w	hich are not searchable fror	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 define allele ids using the drop-do		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 sub	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	♥ =	♥ 5		Display. 23 • records per page ()
	у by	Action		
		RESET	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 TCGTCGAACTCGAT		

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 + 0 locus v ascending v	
allele id v = v 5 Display: 25 v records per page 🕥	
Filter guery by Action	
RESET SEARCH	
1 record returned. Click the hyperlink for detailed information.	
locus     allele id     sequence     gequence length     gequence     gequence	

BIGSdb		
Home > Organisms > Organis	m > Neisseria typing > Allele information	
Allele informa	tion - abcZ: 5	Help 🔽
Provenance/me	eta data	
locus:	abcZ	
allele:	5	
sequence:	TTIGATACCG TIGCCGAAGG TTIGGGCGAA ATTCGCGATT TATTGCGCCG TTATCATCAT GTCAGCCATG AGTTGGAAAA TGGTTCGAGT GAGGCTTTGT TGAAAGAGCT TAACGAATTG CAACTTGAAA TCGGAAGCGAA GGACGGCTGG AAGCTGGATG CGGCAGTCAA GCAGACTTG GGTGAACTTG GTTTGCCAGA AAACGAAAAA ATCGGCAACC TCTCCCGGCG ACAGAAAAAG CGTGTTGCCC TAGCGCAGGC TTGGGTGCAG AAGCCTGATG TATTGCTGCT GGACGAACCG ACCAACCATT TGGACATTGA CGCGATTATT TGGCCGGAAA ATCTGCTTAA AGCGTTTGAA GGCAGCCTGG TTGTGGATTAC CCACGACCCG CGTTTTTGG ACCAATATCGC CACGCGCATC GTCGAACTCG ATC	
length:	433	
status:	Sanger trace checked	
date entered:		
	2009-11-11	
	Keith Jolley, University of Oxford, UK	
	Man-Suen Chan, University of Oxford	
📰 Profiles contai	ning this allele	
MLST:	302 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 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 posample. 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 < v 10 Order by: locus v ascending v Display: 25 v records per page ()
Filter query by Action
status:
type allele:
sender:
allele flag:
9 records returned. Click the hyperlinks for detailed information.

locus	allele id	sequence	sequence length	type allele comments fl	ags
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.

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	+
moures.			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.	MALYSIS	+
				+
		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 attribut	e search (locus-specific)	
Sequence attribute search - PorA_	VR2	Help 🔀 🕄 🗡
Locus: PorA_VR2 Page will reload w • Further information is available for this locus. Please enter your search criteria below (or leave blank and submit to re		
Allele fields	Display	Action
allele id 🔹 👻	+ (i) Order by: allele id v ascending v Display: 25 v records per page (j)	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	Hetp 🕜 🚺 🗡
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 Action	n
family     =      2     +     ()     Order by:     ascending ~     RES       Display:     25 ~     records per page ()	SET

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

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Se	quence attribute search (locus-specific)	
Sequence attribute search -	abcZ	Help 🖍 🚯 🥢
Locus:         abcZ         ✓         Pa           •         Further information is available for this locus.           Please enter your search criteria below (or leave blank ar	ge will reload when changed id submit to return all records).	
Allele fields	Display	Action
allele id 🔹 👻 = 👻	→ ①     Order by: allele id     ✓ ascending ✓       Display:     25 ✓ records per page ①	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.	×
Please enter your search criteria below (or leave blank and submit to return all records).	Modify form parameters Click to add or remove additional query terms:
Allele fields Display	Allele fields
allele id v = v + () Order by: allele id v ascending v Display: 25 v records per page ()	<ul> <li>Allele id list box</li> <li>Filters</li> </ul>

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     Allele id list     Display	
allele id v = v + () 3 4 5 Croter 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 (loce	us-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 record  Allele fields allele id  allele id  fields	ds). Allele id list Display 1 Order by: allele id v ascending v 3 Display: 25 v records per page () 4 SACtion RESET SEARCH	

5 records returned. Click the hyperlinks for detailed information.

locus	allele id	sequence	sequence length	type allele	comments flags
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		

Analysis tools		
🐻 Export:	FASTA	Table
🛃 Analysis:	Locus Ex	cplorer

# 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.

 GSdb	) ganisms > Organism > Neisseria typing				
	<i>isseria</i> typing databas	e			
	Query a sequence	Find alleles	Search for allelic profiles	DICG 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	+
	natches.			EXPORT	+
	Batch sequences Query multiple independent sequences	By locus Select, analyse and download specific	<b>By allelic profile</b> This can include partial matches to find related profiles.		+
	in FASTA format to identify allelic matches.	alleles from a single locus.			+
			In a batch	<b>1</b> INFORMATION	+
			Look up multiple allelic profiles together.	S ISOLATES	

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'.

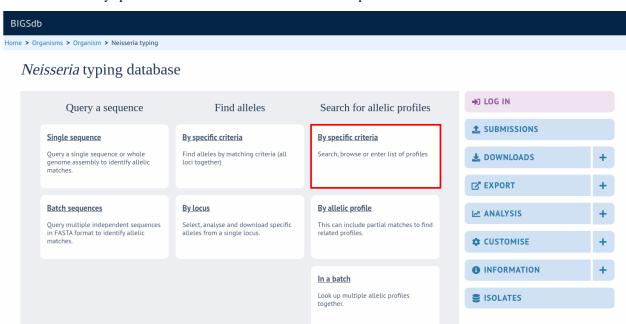
BIGSdb	
Home > Organisms > Organism > Neisseria typing > Search or browse profiles	
Search or browse profiles	Help 🕜 🚯 🖌
Schemes Please select the scheme you would like to query: MLST	
Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.  Locus/scheme fields Display/sort options Action ST v = v Piper to the	]
Browsing all records. 15,577 records returned (1 - 25 displayed). Click the hyperlinks for detailed information. (() () 1 2 3 4 5 6 () ()	
ST       abcZ       abcK       abcK       gdh       pdhC       pgh       clonal complex         1       1       5       1       1       1       5       ST-1 complex         2       1       3       4       7       1       1       3       ST-1 complex         3       1       1       1       3       ST-1 complex       ST-1 complex         4       1       3       3       1       4       2       3       ST-4 complex         5       1       1       2       3       ST-5 complex       ST-5 complex         6       1       2       1       3       2       1       ST-5 complex         7       1       1       2       1       3       2       19       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, Paula Kriz, Anula 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 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  Select		
Enter search criteria or leave blank to browse all records. Modify form parameters to filt 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           13607         8         5         55         10         24         17           13608         3         6         9         17         17         21         9         ST-41/44 complex           13609         5         96         17         15         30         65         12         ST-198 complex           13610         7         5         6         13         82         53         15         ST-213 complex           13611         3         6         34         5         36         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		

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	
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         Combine searches with: AND ×       0rder by: ST × ascending ×       Display: ST × ascending ×       Execution of the searches with: ST × ascending ×       Display: ST × ascending ×       Execution of the searches with: ST	l
134 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.         Image: Constraint of the system	

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

BIGSdb							
lome > Organisms	s > Organism > Ne	eisseria typing > Sea	rch or browse profiles				
Search or	r browse	profiles				Helj	8 1
Schemes Please select the s MLST	scheme you would	like to query: ✓ Select					
Enter search criter		o browse all records.	Modify form parameters to fil	ter or enter a list of values. Display/sort options	Action		

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 > Search or browse prof	iles	
Search or brows	e profiles		Help 🗹 🚺 🗡
Schemes Please select the scheme you we MLST	ould like to query:		
Enter search criteria or leave bla ——Locus/scheme fields——	nk to browse all records. Modify form param	eters to filter or enter a list of values. Display/sort options	★ Modify form parameters
ST • =	v	Order by: ST      ascending      Display: 25      records per page	Click to add or remove additional query terms:
			Attribute values list     Filters

A list box will appear within the page. Hide the form modification tab by clicking the 'X' in the corner or the purple tab again. Now you can choose the attribute to search on along with a list of values.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Search or browse profi	les	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 param Locus/scheme fields ST v = v	Attribute values list Order by: ST V asce Display/sort options Field: ST V Order by: ST V asce Display: 25 V records per page Action RESET SEARCH	
4 records returned. Click the hyperlinks for detailed information.         ST       abc2       adk       aroE       fumC       gdh       pdhC       pgm       clonal complex         1       1       5       1       1       3       ST-1 complex         2       1       3       4       7       1       1       ST-1 complex         3       1       3       1       1       1       23       15       ST-1 complex         4       1       3       1       4       2       3       ST-4 complex		
Analysis tools		

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	ilter or enter a list of values.		×
Locus/scheme fields	Attribute values list	Display/sort optio	Modify form parameters
ST • = • + ()	Field: ST v 1 2 3 4	Order by: ST Display: 25 v Action RESET SE,	Click to add or remove additional query terms:   Click to add or remove additional query terms:

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.

BIGSdl Home > Or	D rganisms > Organism > Neisseria typing				
Ne	<i>isseria</i> typing databas	se			
	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		+
	inderes.			EXPORT	+
	Batch sequences Query multiple independent sequences	By locus Select, analyse and download specific	By allelic profile This can include partial matches to find		+
	in FASTA format to identify allelic matches.	alleles from a single locus.	related profiles.		+
			In a batch	<b>③</b> INFORMATION	+
			Look up multiple allelic profiles together.	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  Select	
Please enter your allelic profile below. Blank loci will be ignored Autofill profile	
abcZ adk aroE fumC gdh pdhC pgm ST: Autofil	
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  Select	
Please enter your allelic profile below. Blank loci will be ignored Autofill profile	
abcZ         adk         aroE         fumC         gdh         pdhC         pgm         ST:         Autofil           2         3         4         3         8         4         6         ST:         Autofil	
Options Display/sort options Action	
Search: Exact or nearest match v Order by: ST v ascending v Display: 25 v records per page ()	

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 📝 🚯
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: 11 Autofill	
Options     Display/sort options     Action       Search:     Exact or nearest match •     Order by: ST • ascending • Display:     RESET     SEARCH	

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	
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 profil	e
abcZ         adk         aroE         fumC         gdh         pdhC         pgm         ST:           2         3         4         3         8         4         6	11 Autofill
Options Display/sort options Action	
Search: Exact or nearest match  Order by: ST ascending Display: 25  records per page ()	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     3     8     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         ST: 11         Autoful           2         3         4         3         8         4         6         5	
Options Display/sort options Action	
Search: 4 or more matches  V Order by: ST V ascending V Display: 25 V records per page ()	
499 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.	
(() () 1 2 3 4 5 6 () ())	
ST         abc2         adk         areE         fumC         gdh         pdhC         pgm         clonal complex           10         2         5         4         2         8         15         2         ST-8 complex           11         2         3         4         3         8         4         6         ST-11 complex           50         2         3         19         3         8         4         6         ST-11 complex           51         2         3         4         2         ST-11 complex         5         5         7         3         4         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	HO 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	Ł 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.		
		In a batch	<b>1</b> 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.

me > Organisms >	Organism 3	Neisseria	typing > I	Batch profile	e query						
atch prof	ile qu	ery									Help 📝
Schemes Please select the scl	neme you wo	ould like to	query:								
	,			_							
MLST			✓ Selee	t							
olumn should be a	n isolate ider										
Enter allelic profiles column should be ar stripped out of the c Paste in profil isolate_1	n isolate ider juery. es 2	ntifier and t	the remaini	ng columns	should com	prise the al	lele numbers (sh				
column should be an stripped out of the c Paste in profil isolate_1 isolate_2	n isolate ider juery. es 2 35	ntifier and t 6 24	the remaini 9 15	ng columns	should com 9 6	prise the al	8 41				
volumn should be an stripped out of the or Paste in profil isolate_1 isolate_2 isolate_3	n isolate ider juery. es 2 35 7	6 24 5	the remaini 9 15 1	ng columns 5 48 13	should com 9 6 82	6 482 53	8 41 17				
column should be an stripped out of the construction Paste in profil isolate_1 isolate_2 isolate_3 isolate_4 isolate_5	n isolate ider juery. es 2 35 7 6 558	6 24 5 5 43	9 15 1 53 39	5 48 13 3 45	9 6 82 74 65	6 482 53 8 594	8 41 17 54 600				
column should be an stripped out of the co- Paste in profil isolate_1 isolate_2 isolate_3 isolate_4 isolate_5 isolate_6	n isolate ider juery. es 2 35 7 6 558 8	6 24 5 5 43 4	9 15 1 53 39 6	5 48 13 3 45 17	9 6 82 74 65 5	6 482 53 8 594 18	8 41 17 54 600 16				
column should be an stripped out of the c Paste in profil isolate_1 isolate_2 isolate_3 isolate_4 isolate_5 isolate_6 isolate_7	n isolate ider juery. 2 35 7 6 558 8 222	6 24 5 5 43 4 3	9 15 1 53 39 6 58	5 48 13 3 45 17 275	9 6 82 74 65 5 30	6 482 53 8 594 18 3	8 41 17 54 600 16 255				
column should be an stripped out of the c Paste in profil isolate_1 isolate_2 isolate_3 isolate_4 isolate_5 isolate_6 isolate_7 isolate 8	n isolate ider juery. 2 35 7 6 558 8 222 222	6 24 5 5 43 4 3 1	9 15 1 53 39 6 58 58 58	5 48 13 3 45 17 275 261	9 6 82 74 65 5 30 263	6 482 53 8 594 18 3 5	8 41 17 54 600 16 255 255				
column should be an stripped out of the c 	n isolate ider juery. 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 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				
column should be an stripped out of the c 	n isolate ider juery. 2 35 7 6 558 8 222 222	6 24 5 5 43 4 3 1	9 15 1 53 39 6 58 58 58	5 48 13 3 45 17 275 261	9 6 82 74 65 5 30 263	6 482 53 8 594 18 3 5	8 41 17 54 600 16 255 255				
column should be an stripped out of the c 	n isolate ider juery. 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 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				
column should be an stripped out of the c 	n isolate ider juery. 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 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.

IGSdb me > Orga	nism	; > 0	rgani	sm >	Neiss	seria t	yping	> Bato	h profile query			
1-		<b>c</b> :1				_					He	lp 🖸
Batch	рго		e	que	егу							
				fumC	gdh	pdhC	pgm			Query		
isolate_1	2	6	9	5	9	6	8	3021	ST-41/44 complex	d		
isolate_2 isolate_3	35	24	15	48 13	82	482 53	41 17	7697	ST-192 complex ST-213 complex	à		
isolate 4	6	5	53	3	74	8	54	9403	ST-215 complex	ã		
		43	39	45	65	594			ST-1494 complex (lactamica)	à		
isolate 6	8	4	6	17	5	18		4399	ST-103 complex	Q		
	222	3	58	275	30	3		8691	ST-4821 complex	Q		
	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		2	6	25	17	655		10137	ST-53 complex	Q		
isolate_13		10	2	5	3	453	9	7008		Q		
isolate_14		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.

BIGSdI	0				
Home > Or	ganisms > Organism > Neisseria typing				
Ne	<i>isseria</i> typing databas	е			
	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	L DOWNLOADS	+
					+
	Batch sequences Ouery multiple independent sequences	By locus Select, analyse and download specific	By allelic profile This can include partial matches to find		-
	in FASTA format to identify allelic matches.	alleles from a single locus.	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	milarity	
Find most similar alleles		Help 🔽
This page allows you to find the most similar sequences to a selected allele u	sing BLAST.	
Select parameters	Action	
Locus: abcZ  Allele: 5 Number of results: 10	RESET	

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

IGSdb												
ome > Orga	inisms > Org	ganism > 1	Veisse	ria typing > F	Plugins > S	equence s	similarity					
ind n	nost si	milaı	al	leles								Help 🔽
This page a <sup>i</sup>	llows you to	find the mo	ost sim	ilar sequence	es to a selec	ed allele	using BLAST					
Select	parameters	s					Actior					
	Loc	us: abcZ ele: 5				~	RESI	UBMIT				
		lismatches		Alignment C								
Allele 1 abcZ: 729	99.769	1	0	433/433	<u>4</u>							
Allele abcZ: 729 abcZ: 671	99.769 99.769	1	0	433/433 433/433	<u>4</u>							
Allele abcZ: 729 abcZ: 671 abcZ: 657	99.769 99.769 99.769	1 1 1	0 0 0	433/433 433/433 433/433	ৰাঁণ্ড ৰাঁণ্ড ৰাঁণ্ড							
Allele : abcZ: 729 abcZ: 671 abcZ: 657 abcZ: 453	99.769 99.769 99.769 99.769	1 1 1 1	0 0 0	433/433 433/433 433/433 433/433	46 46 46 46							
Allele 2 abcZ: 729 abcZ: 671 abcZ: 657 abcZ: 453 abcZ: 213	99.769 99.769 99.769 99.769 99.769	1 1 1 1 1	0 0 0 0	433/433 433/433 433/433 433/433 433/433	46 46 46 46							
Allele abcZ: 729 abcZ: 671 abcZ: 657 abcZ: 453 abcZ: 213 abcZ: 166	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	433/433 433/433 433/433 433/433 433/433 433/433	ଶ୍ ଅ ଶ୍ ଶ୍ ଶ୍ ଶ୍ ଶ୍							
Allele 9 abcZ: 729 abcZ: 671 abcZ: 453 abcZ: 213 abcZ: 166 abcZ: 114	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	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							
Allele abcZ: 729 abcZ: 671 abcZ: 657 abcZ: 453 abcZ: 213 abcZ: 166	99.769       99.769       99.769       99.769       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	0 0 0 0 0 0 0 0 0	433/433 433/433 433/433 433/433 433/433 433/433	ଶ୍ ଅ ଶ୍ ଶ୍ ଶ୍ ଶ୍ ଶ୍							

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	
Locus: abcZ	
Allele #2: 729	
Nucleotide differences between abcZ: 5 and abcZ: 729	
Identity: 99.77 %	
Show alignment	
Differences: 1 157: 6 - 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.		+
			<b>()</b> INFORMATION	+
			SISOLATES	

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

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

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

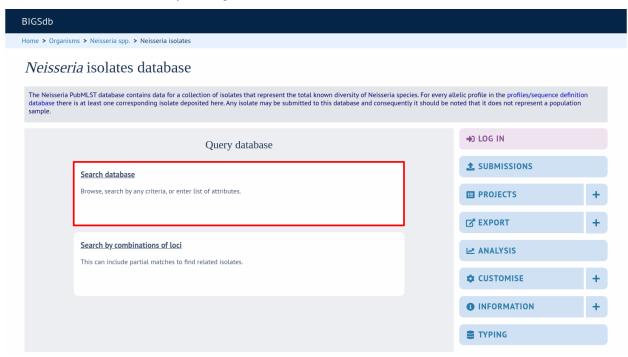
BIGSdb	
Home > Organisms > Organism > Neisseria typing > Plugins > Sequence comparison	
Allele sequence comparison	Help 🔽
This tool allows you to select two alleles and highlight the nucleotide differences between them.       Select parameters     Action       Locus:     abcZ     V       Allele #1:     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: $A \rightarrow G$ 79: $A \rightarrow G$ 79: $A \rightarrow G$ 81: $T \rightarrow C$ 82: $G \rightarrow A$ 83: $G \rightarrow A$ 83: $G \rightarrow A$ 83: $G \rightarrow A$ 89: $G \rightarrow A$	

#### 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'.

lome > Organisr	ms > Neisseria spp. > Neis	sseria isolates	> Search or	browse database							
Search c	or browse dat	tabase								Н	elp 🕜 🚯 .
Enter search cri	iteria or leave blank to brow	vse all records.	Modify form	n parameters to filter or e	enter a list of values.						
Isolate pr	rovenance fields				Display/sort options						
id	× =	✓ Enter va	ue	+ ()	Order by: id				¥ as	cending	*
		Enter va			Display: 25 v re	cords per page	â		· us	contailing	
Action							0				
RESET	SEARCH										
		Click the hores	1								
	returned (1 - 25 displayed).		links for de	tailed information.							
			links for de	tailed information.							
	returned (1 - 25 displayed).	00	$\mathbf{\hat{o}}$					MLST	Finet	vping ant	tioens
	returned (1 - 25 displayed).	00	links for de		species	capsule group	ST	MLST clonal complex		yping anti PorA_VR2	
66,602 records	returned (1 - 25 displayed). 1 2 3 4 5 6	00	Isolate field year 1937 inva	s () disease asive (unspecified/other)	Neisseria meningitidis	A	ST				2 FetA_VR F1-5
66,602 records (()) ((	returned (1 - 25 displayed). 1 2 3 4 5 6 allases	Country USA Pakistan	Isolate field year 1937 inva 1967 mer	s (j) disease asive (unspecified/other) ningitis and septicaemia	Neisseria meningitidis Neisseria meningitidis	A A	1		PorA_VR1 5-2 5-2	PorA_VR2 10 10	2 FetA_VR
66,602 records (() [solate 1 A4/M1027 2 120M 3 M00242905	returned (1 - 25 displayed). <b>1 2 3 4 5 6 allases</b> B1; NIBSC_2803; Z1001	country USA	Isolate field year 1937 inva 1967 met 2000 inva	s () disease asive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A B		clonal complex	PorA_VR1 5-2	PorA_VR2 10	2 FetA_VR F1-5

4	M1027	B43; NIBSC_3076; Z1043	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex			
5	M00240227		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	В	1100	ST-32 complex	7	16	
6	M00282207		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	W	1101	ST-22 complex			
7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1
8	M00242007		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	В	1102	ST-18 complex		14	
9	0021/84		Czech Republic	1984	invasive (unspecified/other)	Neisseria meningitidis	W	114	ST-22 complex			
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	18-1	3	F5-1

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

BIGSdb					
lome > Organisms > Neisseria spp. > Neis	seria isolates > Search or brow	se database			
Search or browse dat	abase				Help 🕜 🚺 🗡
Enter search criteria or leave blank to brow	se all records. Modify form para	meters to filter or	renter a list of values.		
Isolate provenance fields			Display/sort options		
id 🕶 =	✓ Enter value	+ ()	Order by: id Display: 25 v records per page (j)		✓ ascending ✓
Action RESET SEARCH					
66,602 records returned (1 - 25 displayed).	Click the hyperlinks for detailed	information.			
(() (i) 1 2 3 4 5 6	$\odot \odot$				
	Isolate fields (i)			MLST	Finetyping antigens
id isolate aliases	country year	disease	species capsule group S	T clonal complex	PorA_VR1 PorA_VR2 FetA_VR

id	isolate	aliases	country	year	disease	species	capsule group	ST	clonal complex	PorA_VR1	PorA_VR2	FetA_VR
1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A			5-2	10	F1-5
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicaemia	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
3	M00242905		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	В	1099		19	15	
4	M1027	B43; NIBSC_3076; Z1043	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex			
5	M00240227		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	В	1100	ST-32 complex	7	16	
6	M00282207		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	W	1101	ST-22 complex			
7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1
8	M00242007		UK	2000	invasive (unspecified/other)	Neisseria meningitidis	В	1102	ST-18 complex		14	
9	0021/84		Czech Republic	1984	invasive (unspecified/other)	Neisseria meningitidis	W	114	ST-22 complex			
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	18-1	3	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									
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 consequen sample.			'n						
Query database	Query database								
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	+						

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 Isolate provenance fields	or enter a list of values. Display/sort options
Combine with: AND -	Order by: id 🔹 v ascending v
country • = • USA + i	Display: 25 v records per page 👔
year v >= v 2000	Action
	RESET SEARCH

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

GSdb													
me > Organisms > Ne	eisseria spp.	> Neis	seria isolates > Search or br	owse database									
earch or br	rowse	dat	abase								He	₩ <b>[</b> ]	8] /
Enter search criteria or l	leave blank	to brow	vse all records. Modify form pa	arameters to filter or e	enter a list of v	alues.							
Isolate provenance	ce fields—				Display/sort	options							
Combine with: AND	~				Order by:	id				<b>∨</b> a	scending ·	~	
country	<b>v</b> =		✓ USA	+ ()			cords per page 😱	)					
year	✓ >=		✓ 2000		Action			,					
					RESET	SEAR	сн						
	(1 - 25 displa 2 <b>3 4</b>		lick the hyperlinks for detaile	L	RESET	SEAR	СН						
			<b>OO</b>	L	RESET	SEAR		Finet	vning antig	ens			
( 1 2		56		L	RESET		CH MLST clonal complex		yping antig PorA_VR2				
id   isolate   alia:     41   M7085	ses country USA	5 6	Isolate fields () disease invasive (unspecified/other)	ed information. species Neisseria meningitidi	capsule grou s W	ир ST 11	MLST clonal complex ST-11 complex						
(x)         (x)         1         2           Id         isolate         alia:           41         M7085         99         MDL01A0601	ses country USA USA	5 6	Isolate fields () disease invasive (unspecified/other) meningitis	d information. species Neisseria meningitidi Neisseria meningitidi	capsule grou s W s Y	10 ST 11 1378	MLST clonal complex ST-11 complex ST-23 complex						
Isolate         alia:           1         Isolate         alia:           1         M7085         99           9         MDL01A0601         00           00         MDL01A2447         1	ses country USA USA USA	5 6	Isolate fields () disease invasive (unspecified/other) meningitis invasive (unspecified/other)	ed information. species Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi	capsule grou s W s Y s Y	ир ST 11 1378 1379	MLST clonal complex ST-11 complex ST-23 complex ST-23 complex						
Image: Solution         Solution         Solution           Image: Solution of Solution         Image: Solution of Solution         Image: Solution of Solution           Image: Solution of Soluti	ses country USA USA USA USA	<b>5 6</b> <b>/ year</b> 2000 2001 2001 2001	Isolate fields () disease invasive (unspecified/other) meningitis invasive (unspecified/other)	d information. species Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi	capsule grou s W s Y s Y s	ир ST 11 1378 1379 1624	MLST clonal complex ST-11 complex ST-23 complex ST-23 complex ST-23 complex						
Image: Second state	ses country USA USA USA USA USA USA	<b>5 6 / year</b> 2000 2001 2001 2001 2001 2001	Isolate fields () disease invasive (unspecified/other) meningitis invasive (unspecified/other)	d information. <u>species</u> Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi	capsule grou s W s Y s Y s s	up ST 11 1378 1379 1624 1625	MLST clonal complex ST-11 complex ST-23 complex ST-25 complex ST-167 complex ST-26 complex						
Image: Weight of the system         Image: Weight of the system <t< td=""><td>2 3 4 ses country USA USA USA USA USA</td><td><b>5 6 / year</b> 2000 2001 2001 2001 2001 2001 2001</td><td>Isolate fields () disease invasive (unspecified/other) meningitis invasive (unspecified/other)</td><td>species Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi</td><td>capsule grou s W s Y s Y s s s</td><td>up ST 11 1378 1379 1624 1625 1626</td><td>MLST clonal complex ST-11 complex ST-23 complex ST-26 complex ST-26 complex ST-26 complex</td><td>PorA_VR1</td><td>PorA_VR2</td><td>FetA_VR</td><td></td><td></td><td></td></t<>	2 3 4 ses country USA USA USA USA USA	<b>5 6 / year</b> 2000 2001 2001 2001 2001 2001 2001	Isolate fields () disease invasive (unspecified/other) meningitis invasive (unspecified/other)	species Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi	capsule grou s W s Y s Y s s s	up ST 11 1378 1379 1624 1625 1626	MLST clonal complex ST-11 complex ST-23 complex ST-26 complex ST-26 complex ST-26 complex	PorA_VR1	PorA_VR2	FetA_VR			
Image: Second state	ses country USA USA USA USA USA USA	<b>5 6 / year</b> 2000 2001 2001 2001 2001 2001 2001 200	Isolate fields () disease invasive (unspecified/other) meningitis invasive (unspecified/other)	ed information. Species Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi Neisseria meningitidi	capsule grou s W s Y s S s s s s W	up ST 11 1378 1379 1624 1625 1626	MLST clonal complex ST-11 complex ST-23 complex ST-25 complex ST-167 complex ST-26 complex						

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.

BIGSdb					
Home > Organisms > Neisseria spp. > Nei	sseria isolates > Search or browse database				
Search or browse dat	abase			Help	23/
Enter search criteria or leave blank to brow	vse all records. Modify form parameters to filter or e	enter a list of values	i.		
Isolate provenance fields		Display/sort optic			
Combine with: AND ~		Order by: id		✓ ascending ✓	
country v =	• USA + ()	-			
	<ul> <li>✓ 2000</li> </ul>	Display: 25	🖌 records per page 👔		
year 🗸 >=	- 2000	Action			
2.682 records returned (1 - 25 direct-wed) (	lick the hyperlinks for detailed information.	RESET SE	ARCH		
(a) (c) 1 2 3 4 5 6	Isolate fields ()		MLST Finetyr	ping antigens	
id isolate aliases country year	disease species	capsule group	ST   clonal complex   PorA_VR1   P		
341 M7085 USA 2000	invasive (unspecified/other) Neisseria meningitidis		11 ST-11 complex		
499 MDL01A0601 USA 2001	meningitis Neisseria meningitidi		378 ST-23 complex		
	invasive (unspecified/other) Neisseria meningitidis		379 ST-23 complex		
866 MD01227 USA 2001	Neisseria meningitidi		524 ST-167 complex		
867 MD01056 USA 2001	Neisseria meningitidi		525 ST-23 complex		
868 MD01066 USA 2001	Neisseria meningitidi		526 ST-269 complex	2	
	invasive (unspecified/other) Neisseria meningitidi		11 ST-11 complex 5	2 F1-1	
	invasive (unspecified/other) Neisseria meningitidis invasive (unspecified/other) Neisseria meningitidis		ST-11 complex     5       22     ST-22 complex     6	2	
2310 MI/060 USA 2000	invasive (unspecified/other) Neisseria meningitidi	2 11	22 ST-22 complex 6	5	

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

			to brow	vse all records. Modify form p	arameters to filter or er	iter a list of va	lues.			×		
	Isolate provenance	e fields —			C	isplay/sort o	ptions-			Modi	fv form	a parameters
Co	mbine with: AND ~					Order by: jo	i				-)	F
	country	<b>v</b> =		✓ USA	+ ()			cords per page 👔		Click to	add or rem	ove additional query terms:
у	/ear	♥ >=		✓ 2000	A	RESET	SEAR			0	Provenance Secondary Allele desig	
										0	Allele desi	gnation counts
										0	Allele desi	gnation status
2,682	2 records returned (1	- 25 displ	ayed). C	lick the hyperlinks for detaile	d information.					0	Tagged seq	uence counts
$\sim$	$\sim$			00						0	Tagged seq	uence status
$(\ll)$	)(<) 1 2	34	56	()(»)							Attribute v	alues list
	0									O	Filters	
				Isolate fields 🕢				MLST	Fine			
				disease	species	capsule group			orA V/P1			
id	isolate alias	es country	year	disease	species	capsule group	ST	clonal complex Po				
	M7085	USA		invasive (unspecified/other)			5 ST 11	ST-11 complex				
341 499	M7085 MDL01A0601	USA USA	2000 2001	invasive (unspecified/other) meningitis	Neisseria meningitidis Neisseria meningitidis	W	11 1378	ST-11 complex ST-23 complex				
341 499 500	M7085 MDL01A0601 MDL01A2447	USA USA USA	2000 2001 2001	invasive (unspecified/other) meningitis invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W	11 1378 1379	ST-11 complex ST-23 complex ST-23 complex				
341 499 500 866	M7085 MDL01A0601 MDL01A2447 MD01227	USA USA USA USA	2000 2001 2001 2001	invasive (unspecified/other) meningitis invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	W Y Y	11 1378 1379 1624	ST-11 complex ST-23 complex ST-23 complex ST-167 complex				
341 499 500 866 867	M7085 MDL01A0601 MDL01A2447 MD01227 MD01056	USA USA USA USA USA	2000 2001 2001 2001 2001	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	ST-11 complex ST-23 complex ST-23 complex ST-167 complex ST-23 complex				
341 499 500 866 867 868	M7085 MDL01A0601 MDL01A2447 MD01227 MD01056 MD01066	USA USA USA USA USA USA	2000 2001 2001 2001 2001 2001	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	ST-11 complex ST-23 complex ST-23 complex ST-167 complex ST-23 complex ST-269 complex				
341 499 500 866 867 868 2281	M7085 MDL01A0601 MDL01A2447 MD01227 MD01056 MD01066 M7089	USA USA USA USA USA USA	2000 2001 2001 2001 2001 2001 2001 2000	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	ST-11 complex ST-23 complex ST-23 complex ST-167 complex ST-23 complex ST-269 complex ST-11 complex	5	2	F1-1	
341 499 500 866 867	M7085 MDL01A0601 MDL01A2447 MD01227 MD01056 MD01066	USA USA USA USA USA USA	2000 2001 2001 2001 2001 2001 2000 2000	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	ST-11 complex ST-23 complex ST-23 complex ST-167 complex ST-23 complex ST-269 complex		2 2 3	F1-1	

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.

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 designations/scheme fields  Id  V = V Enter value  Display/sort options  Order by: id  Display: 25 V records per page ()  RESET SEARCH	<ul> <li>Modify form parameters</li> <li>Click to add or remove additional query terms:</li> <li>Provenance fields</li> <li>Secondary metadata</li> <li>Allele designations/scheme field values</li> <li>Allele designation status</li> <li>Allele designation status</li> <li>Tagged sequence counts</li> <li>Tagged sequence status</li> <li>Attribute values list</li> <li>Filters</li> </ul>

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.

ome > Organisms >	<ul> <li>Neisseria spp.</li> </ul>	<ul> <li>Neisser</li> </ul>	ia isolates > Search or	browse database					
Search or	browse	data	base					Help 🕜	9 /
Enter search criteria	a or leave blank	to browse	all records. Modify form	parameters to filter or e	nter a list of val	ues.			
Isolate prover	nance fields				Allele designat	tions/scheme fields -			
			. Esta e value		5				
id	✓ =	``	Enter value	i + i s	T (MLST)	=	✓ 11	+ (i)	
——Display/sort o	options				Action				
Order by: j	id			✓ ascending ✓					
				· ascending ·	RESET	SEARCH			
Display: 2	25 🗙 records n	er nage 🔿	<b>`</b>						
	25 🗸 records p ned (1 - 25 displa		) the hyperlinks for detai	iled information.					
6,991 records return		ayed). Click	the hyperlinks for detai	iled information.					
6,991 records return	ned (1 - 25 displa 2 3 4	ayed). Click	the hyperlinks for detai			MLST			
6,991 records return	ned (1 - 25 displa 2 3 4 s country	ayed). Click 5 6 year	the hyperlinks for detai	species	capsule group	MLST ST clonal complex			
6,991 records return (() () 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ned (1 - 25 displa 2 3 4 s <u>country</u> USA	ayed). Click 5 6 year 1993 inva	the hyperlinks for detai	species Neisseria meningitidis	capsule group C	MLST ST   clonal complex 11 ST-11 complex			
6,991 records return (() () () 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ned (1 - 25 displa 2 3 4 s country USA USA	ayed). Click 5 6 year 1993 inva 1994 inva	the hyperlinks for detait	species	capsule group C	MLST ST clonal complex			
6,991 records return (() () () 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ned (1 - 25 displa 2 3 4 s country USA UK [ScotLand]	ayed). Click 5 6 Year 1993 inva 1994 inva 2000 inva	the hyperlinks for detai	species Neisseria meningitidis Neisseria meningitidis	capsule group C C C	MLST ST clonal complex 11 ST-11 complex 11 ST-11 complex			
6,991 records return (() () () () () () () () () () () () ()	ed (1 - 25 displa 2 3 4 s country USA UK [Scotland] UK [Scotland]	ayed). Click 5 6 1993 inva 1994 inva 2000 inva 2000 inva	the hyperlinks for detain solate fields () disease sive (unspecified/other) sive (unspecified/other) sive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis	Capsule group C C C C C	MLST ST clonal complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex			
6,991 records return (() () () () () () () () () () () () ()	ed (1 - 25 displa 2 3 4 5 country USA USA UK [Scotland] UK [Scotland] UK [Scotland]	ayed). Click 5 6 1993 inva 1994 inva 2000 inva 2000 inva 2000 inva	the hyperlinks for detait solate fields () disease disease disease unspecified/other) sisive (unspecified/other) sisive (unspecified/other) sisive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group C C C C C C C	MLST ST clonal complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex			
6,991 records return (() () () () () () () () () () () () ()	ed (1 - 25 displa 2 3 4 5 country USA USA UK [Scotland] UK [Scotland] UK [Scotland] UK	ayed). Click 5 6 1993 inva 1994 inva 2000 inva 2000 inva 2000 inva 2000 inva	the hyperlinks for detail () () () () () () () () () ()	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	Capsule group C C C C C C C	MLST ST clonal complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex			
6,991 records return (() () () 1 (d) isolate allases 79 160	2 3 4 2 3 4 5 country USA USA USA UK [Scotland] UK [Scotland] UK [Scotland] UK [Scotland]	ayed). Click 5 6 year 1993 inva 1994 inva 2000 inva 2000 inva 2000 inva 2000 inva 2000 inva	the hyperlinks for detait solate fields () disease sive (unspecified/other) sive (unspecified/other) sive (unspecified/other) sive (unspecified/other) sive (unspecified/other) sive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group C C C C C C C C C	MLST ST clonal complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex 11 ST-11 complex			
6,991 records return (() () () () () () () () () () () () ()	ed (1 - 25 displa 2 3 4 5 country USA USA UK [Scotland] UK [Scotland] UK [Scotland] UK [Scotland]	ayed). Click 5 6 1993 invz 1994 invz 2000 invz 2000 invz 2000 invz 2000 invz 2000 invz 2000 invz 2000 invz	the hyperlinks for detait () () () () () () () () () ()	species Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	capsule group C C C C C C C C C C B	MLST ST (clonal complex 11 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	
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 •	Click to add or remove additional query terms:
Action	Provenance fields
RESET	<ul> <li>Secondary metadata</li> <li>Allele designations/scheme field values</li> </ul>
KESET SEARCH	Allele designation sounts
	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'.

ne	<ul> <li>Organisi</li> </ul>	ms > Neisseria spp. > Neiss	eria isolates >	Search	or browse database							Help 🔽 🚺
22	arch c	or browse data	abase									
nte	er search cri	teria or leave blank to brows	e all records. №	1odify for	rm parameters to filter or e	nter a list of values.						
	Isolate pr	ovenance fields				Allele designation cou	nts					
		✓ =						>•		-		
Ľ	id	♥ =	✓ Enter valu	ie	+ (j) Co	ount of total designation	IS	>•	1000 +	<b>(i</b> )		
	Display/s	ort options				Action						
	Order b	ar id			✓ ascending ✓							
		iy: 25 🖌 records per page (	<u>^</u>		· ascending ·	RESET SEA	ARCH					
	Displa	iy. 25 • records per page	(i)									
			0									
			0									
			0									
7,5	28 records	returned (1 - 25 displayed). Cl	0	nks for d	detailed information.							
7,5	28 records	, ,	0	nks for d	detailed information.							
,5 ≪	28 records	returned (1 - 25 displayed). Cl 1 2 3 4 5 6	0	nks for d	detailed information.							
,5 «	28 records	, ,	0	nks for d	detailed information.							
,5 «	28 records	, ,	lick the hyperli	nks for d					MLST	Finet	vping antic	gens
,5 «	28 records	, ,	lick the hyperli	)		species	capsule gro	up ST	MLST clonal complex		yping antig PorA_VR2	
×	)()	1 2 3 4 5 6	lick the hyperli	solate fie	elds (j)		capsule grou A	up ST				
×	isolate A4/M1027 120M	1 2 3 4 5 6	ick the hyperli	solate fie year 1937 in 1967 n	elds ) disease nvasive (unspecified/other) neningitis and septicaemia	Neisseria meningitidis Neisseria meningitidis		1	clonal complex ST-1 complex	PorA_VR1 5-2 5-2	PorA_VR2	FetA_VR F1-5 F5-1
×	isolate A4/M1027 120M 7891	1 2 3 4 5 6 aliases B1; NIBSC 2803; Z1001 B35; NIBSC 2823; Z1035 B54; NIBSC 2760; Z1054	ick the hyperli () () () () () () () () () ()	solate fie year 1937 in 1967 n 1975 in	elds ) disease nvasive (unspecified/other) nvasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A	1	Clonal complex 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
×	isolate A4/M1027 120M 7891 6748	1 2 3 4 5 6	iick the hyperli Country USA Pakistan Finland Canada	solate fie year 1937 in 1967 n 1975 in 1971 in	disease nvasive (unspecified/other) meningitis and septicaemia nvasive (unspecified/other) nvasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A	1 5 1	Clonal complex 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
×	isolate A4/M1027 120M 7891 6748 129E	1         2         3         4         5         6           B1; NIBSC 2803; 21001         B35; NIBSC 2822; 21035         B35; NIBSC 2760; 21054         B73; NIBSC 2760; 21054           B73; NIBSC 2784; 21073         B92; NIBSC 2784; 21092         B92; NIBSC 2784; 21092         B92; NIBSC 2784; 21092	ick the hyperli Country USA Pakistan Finland Canada Germany	solate fie year 1937 in 1967 n 1975 in 1971 in 1964 in	elds ) disease nvasive (unspecified/other) nvasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A	1 5 1	Clonal complex ST-1 complex ST-5 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
e e e e e e e e e e e e e e e e e e e	isolate A4/M1027 120M 7891 6748 129E 139M	allases           B1; NIBSC 2803; Z1001           B35; NIBSC 2822; Z1035           B54; NIBSC 2760; Z1054           B73; NIBSC 2784; Z1073           B92; NIBSC 2784; Z1073           B92; NIBSC 2785; Z1099	ick the hyperli Country USA Pakistan Finland Canada Germany Philippines	solate fie year 1937 in 1967 n 1975 in 1971 in 1964 in 1968	disease disease nvasive (unspecified/other) meningitis and septicaemia nvasive (unspecified/other) nvasive (unspecified/other) nvasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A A A	1 5 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	PorA_VR2 10 10 9 3 10 10	FetA_VR F1-5 F5-1 F3-1 F5-1 F3-6 F5-1
«	isolate A4/M1027 120M 7891 6748 129E	1         2         3         4         5         6           B1; NIBSC 2803; 21001         B35; NIBSC 2822; 21035         B35; NIBSC 2760; 21054         B73; NIBSC 2760; 21054           B73; NIBSC 2784; 21073         B92; NIBSC 2784; 21092         B92; NIBSC 2784; 21092         B92; NIBSC 2784; 21092	ick the hyperli Country USA Pakistan Finland Canada Germany	solate fie year 1937 in 1967 n 1975 in 1971 in 1974 in 1968 1973 in	disease nvasive (unspecified/other) meningitis and septicaemia nvasive (unspecified/other) nvasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A	1 5 1 1 1 4	Clonal complex ST-1 complex ST-5 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 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 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 counts	
id • = • Enter value + () Count of abcZ >• 1 + ()	
Display/sort options Action	
Order by: id vascending V Display: 25 vrecords per page	
10 records returned. Click the hyperlinks for detailed information.	

				Isolate	fields 👔				MLST	F	inetyping ant	igens
id	isolate	aliases	country	year	disease	species	capsule group	ST	clonal complex	PorA_V	R1 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

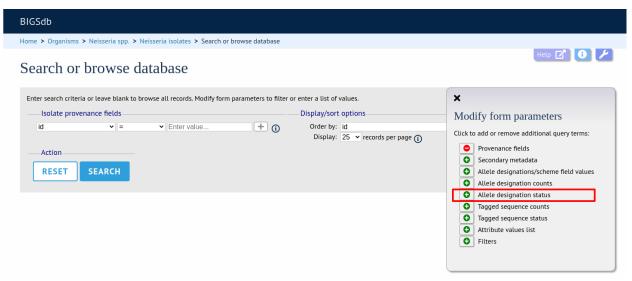
Additional search terms can be combined using the '+' button. Designation count queries will be combined with terms entered in other sections.

**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	
Display: 25 ▼ records per page ()	
2 records returned. Click the hyperlinks for detailed information.           Isolate         Isolate fields ()         MLST         Finetyping antigens           Id         Isolate         aliases         country year disease         species         capsule group         ST         clonal complex         PorA_VR1         PorA_VR2         FetA_VR           14733         N9703         BennettTree32         Unknown         Neisseria subflava         10304	
19363         961-5945         Unknown         Neisseria meningitidis         B         153         ST-8 complex         21         16         F5-8	
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 • = • Enter value + • Display/sort options Ac Order by: id Display: 25 • records per page • Note: Second Secon	Provenance fields

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

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 Isolate provenance fields Id	er a list of values. Annotation status Combine with: AND • Ribosomal MLST • good • + i MLST • good •
Display/sort options Order by: id Display: 25 records per page 3	Action ascending RESET SEARCH
31,286 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.	

$\odot$	1	2	3	4	5	6	$\odot$
---------	---	---	---	---	---	---	---------

			ls	olate f	ields í				MLST	Finet	yping antig	jens
id	isolate	aliases	country	year	disease	species	capsule group	ST	clonal complex	PorA_VR1	PorA_VR2	FetA_VR
1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	5-2	10	F1-5
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicaemia	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-1
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	А	4	ST-4 complex	7	13-1	F1-5

### 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, the number of contigs, the N50 and/or the L50 values.

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 • = • Enter value + • Display/sort options Order by: id Display: 25 • records per page •	<ul> <li>Modify form parameters</li> <li>Click to add or remove additional query terms:</li> <li>Provenance fields</li> <li>Secondary metadata</li> <li>Altele designations/scheme field values</li> <li>Altele designation counts</li> <li>Altele designation status</li> <li>Sequence bin</li> <li>Tagged sequence counts</li> <li>Tagged sequence status</li> <li>Attribute values list</li> <li>Filters</li> </ul>

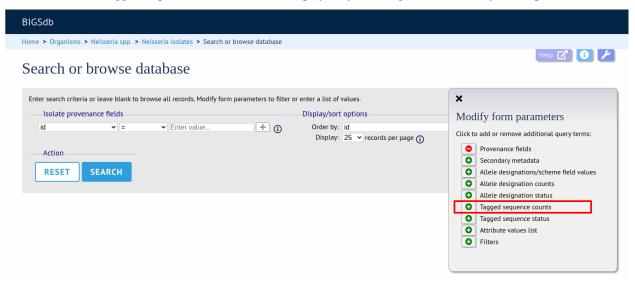
Additional search terms can be combined using the '+' button. Sequence bin queries will be combined with terms entered in other sections.

IGSd	lb											
ome >	Organisms >	• Neisseria spp. > Neisseria isolates > Search or brows	e database									
		1 1.1								Help		
ear	rch or	browse database										
Enter s	search criteria	or leave blank to browse all records. Modify form param	eters to filter or enter	a list of	values.							
	solate prover		Γ		ence bin							
id		▼ = ▼ Enter value	+ (1)		with: AND •							
				total	length (Mbp) 🔹 > 🔹	1.8 🗘 🕂	<b>(i)</b>					
			L	numb	er of contigs 🗐 <= 🔹							
D	isplay/sort o					Action						
	Order by:				<ul> <li>ascending</li> </ul>	RESET S	EARCH					
573 re	Display:	a records per page ① d (1 - 25 displayed). Click the hyperlinks for detailed info	rmation.		▼ ascending ▼	RESET	EARCH					
573 re	Display:	25 records per page (1) d (1 - 25 displayed). Click the hyperlinks for detailed info 2 3 4 5 6 (2) (2)	ormation. olate fields ()		<ul> <li>ascending</li> </ul>	RESET	EARCH		MLST	Fine	typing antiq	gens
573 re (()	Display:	25 records per page (1) d (1 - 25 displayed). Click the hyperlinks for detailed info 2 3 4 5 6 (2) (2)		year	ascending v ascending v disease	RESET S	capsule	ST	MLST clonal complex		typing anti-	<u>г</u>
(() ()	Display:	25 • records per page () d (1 - 25 displayed). Click the hyperlinks for detailed info 2 3 4 5 6 () ())	olate fields 🛈	year 1999			capsule	ST 53				FetA
(() id	Display:	25 records per page (1) d (1 - 25 displayed). Click the hyperlinks for detailed info 2 3 4 5 6 (2) (2) 15 15 15 15 15 15 15 15	olate fields (j)		disease	species Neisseria	capsule group		clonal complex	PorA_VR:	PorA_VR2	FetA F5
id 30	Display: cords returne () 1 isolate 14	25 records per page (1) d (1 - 25 displayed). Click the hyperlinks for detailed info 2 3 4 5 6 (2) (2) Ist BennettTree10; alpha14	olate fields () country Germany	1999	disease carrier invasive	species Neisseria Neisseria	capsule group cnl	53	clonal complex ST-53 complex	PorA_VR: 7	PorA_VR2 30-3	FetA F5 F3
id 30 237	Display: 2 cords returne isolate 14 H44/76	25 records per page () d (1 - 25 displayed). Click the hyperlinks for detailed info 2 3 4 5 6 () () Is Is Is BennetTree10; alpha14 44/76-3; NIBSC_2724; Z5842	olate fields ① country Germany Norway	1999 1976	disease carrier invasive	species Neiseria meningitidis Neiseria meningitidis Neiseria	capsule group cnl B	53 32	clonal complex ST-53 complex ST-32 complex	PorA_VR: 7 7	PorA_VR2 30-3 16	FetA F5 F3 F1
id 30 237 240	Display: 1 cords returne of 1 isolate 14 H44/76 MC58	25 records per page () d (1 - 25 displayed). Click the hyperlinks for detailed info 2 3 4 5 6 () () () solutions	olate fields ① country Germany Norway UK	1999 1976 1983	disease carrier invasive (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis Neisseria Neisseria	capsule group cnl B B	53 32 74	clonal complex ST-53 complex ST-32 complex ST-32 complex	PorA_VR: 7 7 7 7	PorA_VR2 30-3 16 16-2	

#### 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.



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

BIG	Sdb												
Hom	Home > Organisms > Neisseria spp. > Neisseria isolates > Search or browse database												
Search or browse database										9 1			
En	Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.												
	Isolate nr	ovenance fields				Tagged sequence coun	ts						
	id	v =	✓ Enter valu			ount of total tags		> •	1000 +	0			
	lu	• =		ie	+ (j) Co	ount of ioial lags		<b>&gt;</b> ¥	1000	<b>(i)</b>			
	—Display/s	ort options				Action							
	Order t Displa	oy: id ay: 25 v records per page (	1		✓ ascending ✓	RESET SEA	ARCH						
27,	521 records	returned (1 - 25 displayed). Cl	lick the hyperli	nks for	detailed information.								
(	$\otimes$	1 2 3 4 5 6	$\odot$	)									
			1	solate f	fields (i)				MLST	Finet	yping anti	gens	
id	isolate	aliases	country	year	disease	species	capsule gro	up ST	clonal complex				
	A4/M1027	B1; NIBSC_2803; Z1001	USA		invasive (unspecified/other)		Α			5-2	10	F1-5	
2	120M	B35; NIBSC_2822; Z1035	Pakistan		meningitis and septicaemia		Α		ST-1 complex	5-2	10	F5-1	
7	7891	B54; NIBSC_2760; Z1054	Finland		invasive (unspecified/other)		A		ST-5 complex	20	9	F3-1	
10	6748	B73; NIBSC_2784; Z1073	Canada		invasive (unspecified/other)		A		ST-1 complex	18-1	3	F5-1	
11 13	129E 139M	B92; NIBSC_2828; Z1092 B99; NIBSC_2795; Z1099	Germany Philippines		invasive (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A		ST-1 complex ST-1 complex	5-2 5-2	10 10	F3-6 F5-1	
19	S3131	B213; NIBSC 2813; Z1213	Ghana		invasive (unspecified/other)		A		ST-4 complex	5-2	13-1	F1-5	
24	S4355	B213; NIBSC_2813; 21213 B227; NIBSC 2806; Z1227	Denmark		invasive (unspecified/other)		A		ST-5 complex	5-1	9	F3-1	
30	14	BennettTree10; alpha14	Germany	1999	carrier	Neisseria meningitidis	cnl		ST-53 complex	7	30-3	F5-5	

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	• > Organisi	ms > Neisseria spp. > Neiss	eria isolates >	Search or bro	wse database							
Se	earch or browse database									Help 🕜 🚯		
Ent	er search cri	teria or leave blank to brows	e all records. M	odify form pa	rameters to filter or e	nter a list of values.						
	Isolate pr	ovenance fields				Tagged sequence cour	nts					
	id	<b>v</b> =	✓ Enter valu	0		ount of BACT000065 (r		× 2	-	<b>i</b>		
			·				pino) –	. 4	· · · · ·	0		
	Display/s	ort options				Action						
	Order b Displa	y: 25  ✓ records per page (	ì		✓ ascending ✓	RESET SE	ARCH					
27.1	187 records	returned (1 - 25 displayed). Cl	ick the hyperlir	iks for detaile	d information.							
27,1	187 records	returned (1 - 25 displayed). Cl	<b>@</b> @	)								
(«	0	1 2 3 4 5 6	() Is	olate fields (i		species	capsula group	ST	MLST		typing antig	
(«	solate	1 2 3 4 5 6 aliases	<b>@</b> @	olate fields (i year	) disease	species Neisseria meninaitidis		o ST	MLST clonal complex			
(« id	0	1 2 3 4 5 6	ls country USA	olate fields ( year 1937 invasiv	) disease re (unspecified/other)	species Neisseria meningitidis Neisseria meningitidis	А			PorA_VR1	PorA_VR2	FetA_VR
(«	isolate A4/M1027	1 2 3 4 5 6 aliases B1; NIBSC_2803; Z1001	Soundary USA Pakistan	olate fields (i year 1937 invasiv 1967 menin	) disease re (unspecified/other) gitis and septicaemia	Neisseria meningitidis	A	1	clonal complex	PorA_VR1 5-2	PorA_VR2 10	FetA_VR F1-5
id 1 2 7	isolate A4/M1027 120M 7891 129E	1 2 3 4 5 6	Souther States S	olate fields ( year 1937 invasiv 1967 menin 1975 invasiv	disease disease re (unspecified/other) gitis and septicaemia e (unspecified/other)	Neisseria meningitidis Neisseria meningitidis	A A A	1	clonal complex ST-1 complex	PorA_VR1 5-2 5-2 20 5-2	PorA_VR2 10 10 9 10	FetA_VR F1-5 F5-1 F3-1 F3-6
id 1 2 7 11	isolate A4/M1027 120M 7891 129E 139M	1         2         3         4         5         6           B1; NIBSC 2803; 21001         B55; NIBSC 2822; 21035         B54; NIBSC 2760; 21054         B92; NIBSC 2760; 21054           B92; NIBSC 2785; 21099         B99; NIBSC 27795; 21099         B99; NIBSC 2795; 21099         B99; NIBSC 2795; 21099	Souther States S	olate fields ( year   1937 invasiv 1967 menin 1975 invasiv 1964 invasiv 1968	) disease ie (unspecified/other) gitis and septicaemia re (unspecified/other) e (unspecified/other)	<ul> <li>Neisseria meningitidis</li> <li>Neisseria meningitidis</li> <li>Neisseria meningitidis</li> <li>Neisseria meningitidis</li> <li>Neisseria meningitidis</li> </ul>	A A A A A	1 5 1 1	Clonal complex ST-1 complex ST-5 complex ST-1 complex ST-1 complex	PorA_VR1 5-2 5-2 20 5-2 5-2 5-2	PorA_VR2 10 10 9 10 10	FetA_VR F1-5 F5-1 F3-1 F3-6 F5-1
id 1 2 7 11 13 19	isolate A4/M1027 120M 7891 129E 139M S3131	allases           B1; NIBSC 2803; Z1001           B35; NIBSC 2823; Z1035           B35; NIBSC 2822; Z1035           B34; NIBSC 2828; Z1092           B99; NIBSC 2785; Z1054           B9213; NIBSC 2813; Z1213	Source of the second se	olate fields (j year 1937 invasiv 1937 invasiv 1967 menin 1975 invasiv 1964 invasiv 1968 1973 invasiv	disease e (unspecified/other) gitis and septicaemia e (unspecified/other) re (unspecified/other) re (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A A	1 5 1 1 4	clonal complex ST-1 complex ST-5 complex ST-1 complex ST-1 complex ST-4 complex	PorA_VR1 5-2 5-2 20 5-2 5-2 5-2 7	PorA_VR2 10 10 9 10 10 10 13-1	FetA_VR F1-5 F5-1 F3-1 F3-6 F5-1 F1-5
id 1 2 7 11 13 19 24	isolate A4/M1027 120M 7891 129E 139M S3131 S4355	1 2 3 4 5 6 B1; NIBSC_2803; Z1001 B35; NIBSC_2802; Z1055 B54; NIBSC_2760; Z1054 B92; NIBSC_2760; Z1054 B92; NIBSC_2765; Z1099 B213; NIBSC_2813; Z1213 B227; NIBSC_2816; Z1227	Is country USA Pakistan Finland Germany Philippines Ghana Denmark	olate fields (i year 1937 invasiv 1967 menin 1975 invasiv 1964 invasiv 1968 1973 invasiv 1974 invasiv	disease re (unspecified/other) gitis and septicaemia re (unspecified/other) re (unspecified/other) re (unspecified/other)	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	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 ST-5 complex	PorA_VR1 5-2 20 5-2 5-2 5-2 7 5-1	PorA_VR2 10 10 9 10 10 10 13-1 9	FetA_VR F1-5 F5-1 F3-6 F5-1 F3-6 F5-1 F1-5 F3-1
(«	isolate A4/M1027 120M 7891 129E 139M S3131	allases           B1; NIBSC 2803; Z1001           B35; NIBSC 2823; Z1035           B35; NIBSC 2822; Z1035           B34; NIBSC 2828; Z1092           B99; NIBSC 2785; Z1054           B9213; NIBSC 2813; Z1213	Is country USA Pakistan Finland Germany Philippines Ghana Denmark Germany	olate fields (j year 1937 invasiv 1967 menin 1975 invasiv 1964 invasiv 1968 1973 invasiv 1974 invasiv 1974 invasiv	) disease re (unspecified/other) glits and septicaemia re (unspecified/other) re (unspecified/other) capreire carrier	Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis Neisseria meningitidis	A A A A A A Cnl	1 5 1 1 4 5 53	clonal complex ST-1 complex ST-5 complex ST-1 complex ST-1 complex ST-4 complex	PorA_VR1 5-2 5-2 20 5-2 5-2 5-2 7	PorA_VR2 10 10 9 10 10 10 13-1	FetA_VR F1-5 F5-1 F3-1 F3-6 F5-1 F1-5

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

**Note:** Searches for 'all loci' with counts that include zero, e.g. 'count of any locus = 0' or with a '<' operator are not supported. This is because such searches have to identify every isolate for which one or more loci are not tagged. In databases with thousands of loci this can be a very expensive database query.

#### 11.10.7 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 id v = v Enter value + () Action RESET SEARCH	Modify form parameters Click to add or remove additional query terms: Provenance fields Secondary metadata Allele designations/scheme field values Allele designation counts
	<ul> <li>Allele designation status</li> <li>Tagged sequence counts</li> <li>Tagged sequence status</li> <li>Attribute values list</li> <li>Filters</li> </ul>

Select a 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.

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 provenance fields Tagged sequence status	
id v = v Enter value + () NEIS0001 (lpxC) is flagged: internal stop codon v + ()	
Display/sort options Action	
Order by: id  Variable ascending  Variable asc	
1 record returned. Click the hyperlink for detailed information.	
1 record returned. Click the hyperlink for detailed information.           Isolate         Isolate fields ①         MLST         Finetyping antigens           Id         isolate         aliases         country         year         disease         species         capsule group         ST         clonal complex         PorA_VR2         FerA_VR2           2077         153         alpha153         Germany         1999         carrier         Neisseria meningitidis         E         60         ST-60         complex         5         2         F1-7	
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_VR2       FetA_VR         2077       153       alpha153       Germany       1999       carrier       Neisseria meningitidis       E       60       ST-60 complex       5       2       F1-7         Analysis tools   <	
1 record returned. Click the hyperlink for detailed information.           Isolate         Isolate fields ①         MLST         Finetyping antigens           Id         isolate         aliases         country         year         disease         species         capsule group         ST         clonal complex         PorA_VR2         FerA_VR2           2077         153         alpha153         Germany         1999         carrier         Neisseria meningitidis         E         60         ST-60         complex         5         2         F1-7	
1 record returned. Click the hyperlink for detailed information.         Isolate fields ()       MLST         Id       Isolate fields ()         Id       Isolate fields ()         Species       capsule group         ST       clonal complex         PorA_VR1       PorA_VR2         FetA_VR       PorA_VR2         Fields       Gorminations         Polymorphic sites       Publications         Sequence bin	

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 id • = • Enter value + ① Display/Sort options Order by: id Display: 25 • records per page ③ Action RESET SEARCH	Modify form parameters Click to add or remove additional query terms: Provenance fields Secondary metadata Altele designations/scheme field values Altele designation counts Altele designation status Tagged sequence counts Tagged sequence status Filters Filters

Select the attribute to query and enter a list of values.

lGSdb															
ome > Organisms >	> Neisseri	a spp. > Ne	isseria iso	olates >	Search or br	owse datab	ase								
Search or	brow	vse da	itaba	se										Hel	• 🕜 🕚 🖌
Enter search criteria —— Isolate proven id	nance fie			ecords. Mo		arameters to	Fie     I	enter a list of val Attribute value eld: id				~			
							2 3 4 5					ĥ			
Display/sort o Order by:	1							Action							
		ords per pa	ge (j			✓ ascend	ling 🗸	RESET	SEARC	Н					
Display:	25 🗸 rec		Ŭ			✓ ascend	ling V	RESET	SEARC	Η					
Display:	25  ✔ rec	nyperlinks fo	or detailed	Isolate	fields (j)						MLST		/ping anti PorA VR2		
Display: 5 records returned. d isolate 4 4//M1027 B1;	25  ✓ rec Click the h aliase NIBSC_280	nyperlinks fo s D3; Z1001	or detailed country USA	Isolate year 1937 inv	fields (j) disea rasive (unspe	nse ecified/othe	r) Neisseri	species la meningitidis	capsule grou A	p ST	clonal complex	PorA_VR1 5-2	PorA_VR2 10	FetA_VR F1-5	
Display: 5 5 records returned. 1 isolate 1 A4/M1027 B1; 2 120M B35;	25  ✓ rec Click the h aliase NIBSC_280	nyperlinks fo s D3; Z1001	or detailed country USA Pakistan	Isolate year 1937 inv 1967 me	fields (i) disea rasive (unspe eningitis and	nse ecified/othe septicaem	r) Neisseri a Neisseri	species ia meningitidis ia meningitidis	capsule grou A A	p ST		PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR	
Display: 5 5 records returned. 1 isolate 4 A4/M1027 B1; 2 120M B35; 5 M00242905	25  ✓ rec Click the h aliase NIBSC_28 ; NIBSC_28	nyperlinks fo 15 03; Z1001 122; Z1035	country USA Pakistan 2 UK	Isolate year 1937 inv 1967 me 2000 inv	fields (j) disea vasive (unspe eningitis and vasive (unspe	ise ecified/othe septicaem ecified/othe	r) Neisseri a Neisseri r) Neisseri	species la meningitidis la meningitidis la meningitidis	capsule grou A	p ST 1 1099	clonal complex ST-1 complex	PorA_VR1 5-2	PorA_VR2 10	FetA_VR F1-5	
Display: 5 records returned. d isolate 1 A4/M1027 B1; 2 120M B35; 5 M00242905	25  ✓ rec Click the h aliase NIBSC_28 ; NIBSC_28	nyperlinks fo s D3; Z1001	country USA Pakistan 2 UK USA	Isolate year 1937 inv 1967 me 2000 inv 1937 inv	fields (i) disea diso disea disea disea disea disea diso diso diso diso diso d	nse ecified/othe septicaem ecified/othe ecified/othe	r) Neisseri a Neisseri r) Neisseri r) Neisseri	species ia meningitidis ia meningitidis	capsule grou A A B	p ST 1 1099 4	clonal complex	PorA_VR1 5-2 5-2	PorA_VR2 10 10	FetA_VR F1-5	
Display: 5 records returned. 4 Isolate 1 A4/M1027 B1; 2 120M B35; 5 M00242905 4 M1027 B43; 5 M00240227 Analysis tool	25 v rec Click the F NIBSC_28 NIBSC_28 ; NIBSC_30 ; NIBSC_30 ]S Fields	nyperlinks fo 15 03; Z1001 122; Z1035	country USA Pakistan 2 UK USA	Isolate year 1937 inv 1967 me 2000 inv 1937 inv 2000 inv	fields (i) disea diso disea disea disea diso disea diso diso diso diso diso di	ise ecified/othe septicaem ecified/othe ecified/othe	r) Neisseri a Neisseri r) Neisseri r) Neisseri	species la meningitidis la meningitidis la meningitidis la meningitidis la meningitidis	capsule grou A A B A B B	p ST 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	
Display: 5 records returned. 4 isolate 1 A4/M1027 B1; 2 120M B35; 5 M00242905 4 M1027 B43; 5 M00240227 Analysis tool	25 v rec Click the h aliase NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_30 S BURST	15 15 103; 21001 122; 21035 176; 21043 Two Field Codons	country USA Pakistan UK USA UK	Isolate year 1937 inv 1967 me 2000 inv 1937 inv 2000 inv	fields () disea rasive (unspe ningitis and rasive (unspe rasive (unspe rasive (unspe	ise scified/othe septicaem scified/othe scified/othe scified/othe	r) Neisseri r) Neisseri r) Neisseri r) Neisseri r) Neisseri <b>ublication</b>	species la meningitidis la meningitidis la meningitidis la meningitidis la meningitidis	apsule grou A B A B	p ST 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	
Display: 5 records returned. 4 Isolate 1 A4/M1027 B1; 2 120M B35; 5 M00242905 4 M1027 B43; 5 M00240227 Analysis tool	25 v rec Click the h aliase NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_28 NIBSC_30 S BURST	15 15 103; 21001 122; 21035 176; 21043 Two Field Codons	country USA Pakistan UK UK UK UK	Isolate year 1937 inv 1937 me 2000 inv 1937 inv 2000 inv 2000 inv	fields () disea asive (unspe eningitis and asive (unspe asive (unspe asive (unspe Polymorphi	ise scified/othe septicaem scified/othe scified/othe scified/othe	r) Neisseri r) Neisseri r) Neisseri r) Neisseri r) Neisseri <b>ublication</b>	species la meningitidis la meningitidis la meningitidis la meningitidis la meningitidis s Sequence b	apsule grou A B A B	p ST 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	

### 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 data	abase	
Search or browse d	atabase		Help 🗹 🗿 🗡
Enter search criteria or leave blank to I	prowse all records. Modify form parameters	s to filter or enter a list of values.	×
Isolate provenance fields		Display/sort options	Modify form parameters
id v = Action RESET SEARCH	Enter value	Order by: id Display: 25   records per page (	

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 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.		PROJECTS	+		
	1	Z EXPORT	+		
Search by combinations of loci This can include partial matches to find related isolates.		MALYSIS			
			+		
			+		

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 > Organi	sms > Neisse	eria spp. > N	eisseria isolat	es > Search	by locus com	binations										
Search	by loc	us cor	nbinat	ions											Help 🗹	
Schemes Please select MLST		ou would like		Select												
Please e	nter your all	elic profile.	below. Blank	loci will be	ignored. —		A	utofill profi	le by sea	arching	emote o	latabas	e			
abcZ	adk	aroE	fumC	gdh	pdhC	pgm		ST:				Auto	ofill			
Filters	·				Option	s										
Proj	ect:			<b>~</b> (i)	Sea	rch: Exact	t or neare	est match 🗸								
🗆 Include	old record ve	rsions			Orde	//sort optic er by: play: 25		s per page (	)	✓ as	cending	~				
Action	SEAF	ксн														

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         5	
Filters Options	
Project: Search: Exact or nearest match •	
Include old record versions     Display/sort options	
Order by: id v ascending v	
Display: 25 v records per page (i)	
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 Autofil	
9         6         9         9         6         9           Filters         Options         Options         Options         Options	
Project: Search: Exact or nearest match -	
□ Include old record versions	
Action 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 🕜 🚯 🛄
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     9     6     9	
Filters Options	
Project: Search: Exact or nearest match V	
Include old record versions     Display/sort options	
Order by: id • ascending •	
Display: 25 v records per page (1)	
Action	
RESET SEARCH	

Click 'Search'.

lgs	db										
ome :	> Organism	s > Neisseria spp. > I	Neisseria isolates	> Search by	locus combinations						
											Help 🔽 🔒 🚼
Sea	rch h	y locus con	mhinatic	nns							
JCu	i ch b	y 10cus coi	monutio	115							
C -l											
SCI	iemes										
Pleas	e select the	scheme you would lik	e to query:								
MLS	т		×	Select							
IVILO	,,			Jelect							
	0		halan Diaski			1	here and the second states of				
		er your allelic profile						remote database			
12	abcZ 9	adk aroE 6 9	fumC 9	gdh 9	pdhC pgm 6 9	ST: 44	4	Autofill			
	Filters	0 0	0	0	Options						
					· ·						
	Project:			<b>~</b> (i)	Search: Exact	or nearest match 🗸					
	Include old	l record versions			<ul> <li>Display/sort optic</li> </ul>	ons					
					Order by:						
					id		<b>∨</b> a	scending 🗸			
					Display: 25	<ul> <li>records per page (i)</li> </ul>					
	Action										
		CE L D CI I									
	RESET	SEARCH									
Exact	t matches fo	und (7 loci).									
195 r	ecords retur	ned (1 - 25 displayed).	Click the hyperlin	nks for detail	ed information.						
$\frown$	$\sim$		. 00								
Ś	$\odot$	1 2 3 4 5		ッ							
	lastat	-11		Isolate fields				MLST	Finet	typing anti	gens
id 41	isolate 19	aliases	country Germany	year 1999	disease carrier	species Neisseria meningitidis		ST clonal complex 44 ST-41/44 complex	POTA_VR1	PorA_VR2	FetA_VR
70	38		Germany	1999	carrier	Neisseria meningitidis		44 ST-41/44 complex			
427	NG E30	NIBSC_2804; Z4692	Norway	1988	carrier	Neisseria meningitidis	В	44 ST-41/44 complex	21	16	F1-7
774	99 182		Canada			Neisseria meningitidis		44 ST-41/44 complex			
792 975	99-132 0069/93		Canada Czech Republic		ive (unspecified/other) carrier	Neisseria meningitidis Neisseria meningitidis		44 ST-41/44 complex 44 ST-41/44 complex	22	14-4	F1-7
1097	3532	Z7184	The Netherlands		currer	Neisseria meningitidis		44 ST-41/44 complex		4	
1603	0213/93		Czech Republic		carrier	Neisseria meningitidis	В	44 ST-41/44 complex	22	14-4	F1-19
1604	0214/93		Czech Republic	1993	carrier	Neisseria meningitidis	В	44 ST-41/44 complex	22	14-4	F1-7
	0244/93		Czech Republic		carrier	Neisseria meningitidis		44 ST-41/44 complex	22	14-4	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'.

BIGSdb	
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       •         continent       • =       • Europe       + ()       Project:       Select options       • ()         year       • >=       • 2014       * MLST profiles:       • ()       • ()         X MLST profiles:       • ()       Sequence bin: Se	
□ Include old record versions Add filter: ✓ Add Display/sort options Action Order by: id ✓ ascending ✓ Display: 25 ✓ records per page ③	
2,302 records returned (1 - 25 displayed). Click the hyperlinks for detailed information. Your projects Select project  Add these records (	
Isolate fields (i) MLST Finetyping antigens	
idisolatealiasescountryyeardiseasespeciescapsule groupSTclonal complexPorA_VR1PorA_VR2FetA_VR230155M14 240001UK [England]2014invasive (unspecified/other)Neisseria meningitidisW11ST-11 complex52F1-130155M14 240002UK [England]2014invasive (unspecified/other)Neisseria meningitidisW11ST-11 complex52F1-130156M14 240007UK [England]2014invasive (unspecified/other)Neisseria meningitidisW11ST-11 complex52F1-130157M14 240001UK [England]2014invasive (unspecified/other)Neisseria meningitidisW11ST-11 complex52F1-130157M14 240013UK [England]2014invasive (unspecified/other)Neisseria meningitidisW11ST-11 complex52F1-1	
30158         M14 240019         UK [England]         2014         invasive (unspecified/other)         Neisseria meningitidis         W         11         ST-11 complex         5         2         F1-1           30159         M14 240021         UK [England]         2014         invasive (unspecified/other)         Neisseria meningitidis         W         8621         ST-11 complex         5         2         F1-1           30160         M14 240026         UK [England]         2014         invasive (unspecified/other)         Neisseria meningitidis         W         11         ST-11 complex         5         2         F1-1           30161         M14 240029         UK [England]         2014         invasive (unspecified/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           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	Neisseria isolates > Search or	browse database			
Search or	browse d	atabase			🚹 Help 🗹	3 🕄 🗡 🔲
Enter search criter	ia or leave blank to b	rowse all records. Modify form	parameters to filter o	r enter a list of values.		×
Isolate prove	enance fields			Display/sort options		Bookmarks
id Action	► =	✓ Enter value	+ ()	Order by: id Display: 25 v 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 species. For every alle database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be no sample.		1
Query database	€ LOG OUT	+
Search database		
Browse, search by any criteria, or enter list of attributes.	A PRIVATE DATA	
Search by combinations of loci	PROJECTS	+
This can include partial matches to find related isolates.	C EXPORT	+
Bookmarks Retrieve dataset from bookmarked queries.		+
		+
	1033	

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  Database configuration  Created  Share  Run query  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	Hetp 🔽
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 ●         X       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.



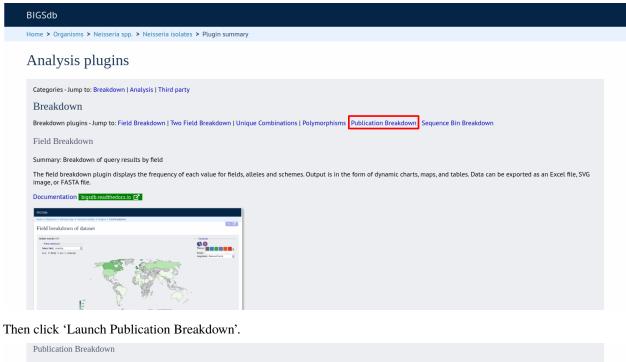
Bookmarks can be deleted by clicking on the delete icon.

### 11.12 Retrieving isolates by linked publication

Click 'Analysis' on the contents page.

BIGSdb			
Home > Organisr	ns > Neisseria spp. > Neisseria isolates		
Neisser	<i>ia</i> isolates database		
	ubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria sp is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequen		
	Query database	+) 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	
			+
			+

Click 'Publication breakdown'



Summary: Breakdown of query results by publication
This 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 to display all isolates linked to any listed publication.
Documentation bigsdbreadthedocsio G*
BGG6 Kern Spanner Normage Network Regis + Materia
Publication breakdown of dataset
- Film way y Duky Artin Aller August 3 Bindly (nameri kana yi asarafa ya Ward Jan ya Jinaya A Jinaya Angala ya Jinaya ya
3 mark diversit 3 3 dispersit da la forbancia diversita. 💿 🔿 1 2 🗿 🌚
Name     Name     Name     Name     Name     Name       1100000000000000000000000000000000000
Launch 'Publication Breakdown'
Sequence Bin Breakdown
Summary: Breakdown of sequence bin contig properties
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 each isolate record. Using this atter metric 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.
Documentation bigsdbreadthedocsio 2*
KGe

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

Home > Organis	sms > Neisseria spp. > Neisseria isolates > Plugins > Publications		
Publica	tion breakdown of dataset		Help 🗹
	ery by Display Action For: All authors Corder by: number of isolates Corder by: SUBMIT Display: 25 Correctors per page SUBMIT		
138 records re	turned (1 - 25 displayed). Click the hyperlinks for detailed information.		
PubMed Year	1 2 3 4 5 6 () ()) r Citation	Title	Isolates in
id Year	r Citation O 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.	Isolates in database 3705 isolates
id <sup>Year</sup> 32163580 2020	r Citation D Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, Tang CM, Maiden MCJ (2020) J Infect Dis [Epub ahead of print]: R Rodrigues CMC, Chan H, Vipond C, Jolley K, Harrison OB, Wheeler J, Whiting G, Feavers IM, Maiden MCJ (2018) Wellcome Open Res 3: 151	Neisseria gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve Surveillance of Antimicrobial	database
id Year 32163580 2020 30687793 2018	r Citation D 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 Open Res 3:151 Rodrigues CMC, Lucidarme J, Borrow R, Smith A, Cameron JC, Moxon ER, Maiden MCJ (2018) Emerg Infect Dis 24: 673- 682	Neisseria gonorrhoeae Population Genomics: Use of the Goncocccal Core Genome to Improve Surveillance of Antimicrobial Resistance. Typing complex meningococcal vaccines to understand diversity and population	database 3705 isolates

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

BIGSdb				
lome > Oi	rganisr	ns > Neisseria spp. > Neisseria isolates > Plugins > Publications		
Publi	cat	ion breakdown of dataset		Help 🗹
		rr Jolley KA v Display Action rr: Jolley KA v Order by: number of isolates v descending v rr: All years v Display: 25 v records per page		
19 record PubMed id	ls retur Year	ned. Click the hyperlinks for detailed information. Clitation	Title	lsolates in database
PubMed id	Year			
PubMed id 32163580	Year 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,	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, Hoffman S, Jolley KA, Kadlubowski M, Kesanopoulos K, Kriz P, Lambertsen L, Levent 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, Izanakaki 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 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				
Home > 0	rganisr	ns > Neisseria spp. > Neisseria isolates > Plugins > Publications		
Publi	cat	ion breakdown of dataset		Help 🗹
Filt		r: Jolley KA v Display Action r: Jolley KA v Order by: number of isolates v descending v r: All years v Display: 25 v records per page SUBMIT		
19 record	ds retur	ned. Click the hyperlinks for detailed information. Citation	Title	Isolates in
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.	Isolates in database 3705 isolates
PubMed id 32163580	Year 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, 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, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos	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, Korsanopoulos K, Kriz P, Lambertsen L, Levenet I, Musilek M, Paragi M, Sauger A, Skoczynska A, Stefanetill P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother <b>51</b> : 2784-92 Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Natl Acad Sci U S A <b>105</b> : 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 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, Kersanopoulos K, Kriz P, Lambertsen L, Levenet I, Musilek M, Paragi M, Sauger A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother <b>51</b> : 2784-92 Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Natl Acad Sci U S A <b>105</b> : 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 <b>15</b> : 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

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

	lb										
Home >	Organisms > N	Neisseria spp. > Neisseria isolates	> Publication	dataset							
Pub	lication	dataset									8
Citati	on query (l	PubMed id: 32163580)									
Cluste	ering with cgMLS	Population Genomics: Use of the ST identified globally distributed, p ions for the emergence of antimics	ersistent, gonod	occal line	ages improving unde	rstanding of the population	biology of gonocod				
3,705	$\sim$	d (1 - 25 displayed). Click the hyper	links for detaile	ed informa	tion.						
3,705	$\sim$		<b>》</b>		tion.			MIST	Einetv	ning ant	igens
3,705 《	$\sim$		<b>》</b>	id informa fields (j	tion. disease	species	capsule group	MLST T clonal con	Finety	ping ant PorA VR2	
«	( 1	2 3 4 5 6 )(	Isolate 1	fields (j)		species Neisseria gonorrhoeae					
K	<ul> <li>1</li> <li>isolate</li> </ul>	2 3 4 5 6 🕥 🤅	Isolate f	fields (i) year 1980			18	T clonal con	nplex PorA_VR1	PorA_VR2	FetA_VR
id 2855 13685 14729	( isolate FA1090 NCCP11945 MS-11	2 3 4 5 6 () allases BennettTree06; NC_002946	Isolate f country USA South Korea USA	fields (i) year 1980		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae	18 19 69	T clonal con 99 01 59	nplex PorA_VR1   18-10 18-10 18-10 18-10	PorA_VR2 43 43 43 43	FetA_VR F1-27 F3-14 F4-7
id 2855 13685 14729 15698	Isolate           FA1090           NCCP11945           MS-11           DGI2	2 3 4 5 6 () () allases BennettTree06; NC_002946 BennettTree05; NC_022240	Isolate f country USA South Korea USA USA	fields () year 1980 2005		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae	18 19 69 84	T clonal con 99 01 59 21	nplex PorA_VR1 1 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 () (2) Construction of the second	Isolate 1 Country USA South Korea USA USA USA	fields () year 1980 2005		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae	18 19 69 84 84	T clonal con 99 01 59 21 18	nplex 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 () () allases BennettTree06; NC_002946 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	18 19 69 84 84 84	T         clonal con           99         01           59         21           18         18	PorA_VR1         PorA_VR1         I           18-10         18-10         18-10           18-10         18-10         18-10           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	Isolate           FA1090           NCCP11945           MS-11           DGI2           PID24	2 3 4 5 6 () (2) Construction of the second	Isolate 1 Country USA South Korea USA USA USA	fields () year 1980 2005		Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae Neisseria gonorrhoeae	18 19 69 84 84 84	T clonal con 99 01 59 21 18	nplex 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           DG18           TCDC-	2 3 4 5 6 () (2) Construction of the second	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	18 19 69 84 84 84 73	T         clonal con           99         01           59         21           18         18	PorA_VR1         I           18-10         18-10           18-10         18-10           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							
Neisseria isolates database							
The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For e database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it shou sample.			n				
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					
			-				
		General options Locus display Scheme display Scheme field display					
		INFORMATION	+				
		S TYPING					

## 11.13.1 General options

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

llGSdb
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 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 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.

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 nucleotides per line in sequence alignments. Display 100 v nucleotides of flanking sequence (where available). Display locus aliases if set. Z Enable toolitips (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 'main results table' tab will scroll up.

GSdb	
me > Organisms > Neisseria spp. > Neisseria isolates > General options	
et database options	3)
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	
Hyperlink allele designations where possible.  Offerentiate 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	

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 ♥] records per page. Page bar position: [top and bottom ♥] Display [100 ♥] nucleotides of flanking sequence alignments. Display [100 ♥] 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	
lome > Organisms > Neisseria spp. > Neisseria isolates > General options	
Set database options	J
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	
House Constant and the second s	
Solate record uspay      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	
<ul> <li>Differentiate provisional allele designations.</li> <li>Display sender, curator and last updated details for allele designations (tooltip).</li> <li>Display information about sequence bin records tagged with locus information (tooltip).</li> <li>Display information about whether alleles have flags defined in sequence definition database (shown in sequence detail tooltip).</li> </ul>	

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.
<ul> <li>◆ General options</li> </ul>
Display 25 v records per page. Page bar position: top and bottom 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 'Provenance field display' tab will scroll up.

BIGSdb			
Home > Organisms > Neisseria spp. > Neisseria i	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			eria isolates) only. If some of the options don't appear to set
General options			
• Main results table - display options			
<ul> <li>Main results table - provenance field selection</li> </ul>			
Select the isolate provenance fields that y performing a locus, scheme or scheme fie dialases strain_designation country continent year date_sampled isoyear_sampled date_received non_culture epidemiological_year age_range ALL NONE DEFAULT	Id query and then selecting the 'Custon age_mth sex Gisease source epidemiology Sepecies Serogroup genogroup_notes Capsule_group MLEE_designation serotype Sero_subtype ET_no penicillin		laying locus and scheme data can be made by  azithromycin bioproject accession private_project sender date_entered date_stamp
<ul> <li>Main results table - secondary metadata selection</li> </ul>	n		
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 an	d 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.

ome > Organisms > Neisseria spp. > Neisseri	a isotates > General options	Het
Set database options		
	ebsite. Options are remembered between sessions and affect th ig 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 s browser.
✓ General options		
Display 25 ▼ records per page. Page bar position: top and bottom ▼ Display 100 < nucleotides per line in s Display 100 ⊂ nucleotides of flankir Display locus aliases if set. 2 Enable tootlips (beginner's mode).		
SET OPTIONS		
• Main results table - display options		
• Main results table - provenance field selection	n	
• Main results table - secondary metadata selec	tion	
Isolate record display		
→ Query filters		
Reset Click the reset button to remove all user setting	gs for this database - this includes locus and scheme field pref	erences.
	gs for this database - this includes locus and scheme field pref	srences.
Click the reset button to remove all user setting		erences.
Click the reset button to remove all user settine		erences.
Click the reset button to remove all user settine <b>RESET ALL TO DEFAULTS</b> Query filters' tab will scroo Query filters	oll up.	erences. uery results. These will be available in the filters section of the query
Click the reset button to remove all user settine RESET ALL TO DEFAULTS Query filters' tab will scroe Query filters Select the fields for which you would like d interface. isolate	DII up.	ery results. These will be available in the filters section of the query
Click the reset button to remove all user settine RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface.  isolate country	oll up.	uery results. These will be available in the filters section of the query
Click the reset button to remove all user settine RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface.  Solotte Country Continent	ropdown lists containing known values on which to filter qu serotype sero_subtype E_no	lery results. These will be available in the filters section of the query           private_project           comments           sender
Click the reset button to remove all user settine 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.	uery 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 settinn RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface. isolate continent region year date_sampled	DII up.	lery results. These will be available in the filters section of the query          private_project         comments         sender         curator         date_entered         date_antpred
Click the reset button to remove all user settine 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 date_sampled isoyear_sampled	DII up.	uery results. These will be available in the filters section of the query  private_project comments sender curator date_entered datestamp WLST profile completion
Click the reset button to remove all user settinn RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface. isolate continent region year date_sampled	DII up.	iery results. These will be available in the filters section of the query
Click the reset button to remove all user settine RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface.  Select the fields for which you would like d interface. Select the fields for whi	DII up.	Jery 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 Conjugative Plasmid 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 ountry continent generic ampled boyear_sampled date_sampled date_received on_culture generic additional and additionand additional and additional and additional and addit	containing known values on which to filter q	uery 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 Cryptic Plasmid profile completion
Click the reset button to remove all user settine RESET ALL TO DEFAULTS Query filters' tab will scree Query filters Select the fields for which you would like d interface.  Select the fields for which you would like d interface. Select the fields for whi	DII up.	uery 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 Cryptic Plasmid profile completion G C ONV peptide typing profile completion N, gonorhoeae cgMLST 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.  Select the fields for which you would like d interface. Select the	DII up.	Jery results. These will be available in the filters section of the query
Click the reset button to remove all user settine 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 date_sampled isoyear_sampled week_sampled lota_caupled lota_cauple	oll up.	Jery 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 Corpite Plasmid profile completion Corpite Plasmid profile completion Corpite Plasmid profile completion Corpite Plasmid profile completion N, gonorrhoeae cgMLST v1.0 profile completion N, meningitdis cgMLST v1.0 profile completion N G MAST profile completion
Click the reset button to remove all user settine RESET ALL TO DEFAULTS  Query filters' tab will scree Query filters  Select the fields for which you would like d interface.  Select the fields for which you would like d i	bill up.         ropdown lists containing known values on which to filter quartering are of subtype         Bero_subtype         Berosubtype         Berosubtype      <	iery results. These will be available in the filters section of the query
Click the reset button to remove all user settine 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 guar date_sampled isoyear_sampled week_sampled lotae_sampled l	DII up.	Jery 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 GC ONUpative Plasmid profile completion Cryptic Plasmid profile completion N, gonorrhoeae cgMLST v1.0 profile completion N G MAST profile completion
Click the reset button to remove all user settine RESET ALL TO DEFAULTS  Query filters' tab will scree Query filters  Select the fields for which you would like d interface.  Select the fields for which you would like d i	bill up.         ropdown lists containing known values on which to filter quartering are of subtype         Bero_subtype         Berosubtype         Berosubtype      <	iery 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 scroe Query filters Select the fields for which you would like d interface. Select the fields for which you would like d select the fields for which you w	copdown lists containing known values on which to filter quarterization in the sero_subtype         sero_subtype         Bero_subtype         Berosubtype         Berosubtype	Jery 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'.

bioproject\_accession
 biosample\_accession
 ENA\_run\_accession

SET OPTIONS

genogroup\_notes capsule\_group MLEE\_designation

NONE

DEFAULT

ALL

## 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 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	+) LC	DG IN						
Search database	<u></u> su							
Browse, search by any criteria, or enter list of attributes.	E PF	ROJECTS	+					
	⊡" E>	(PORT	+					
Search by combinations of loci This can include partial matches to find related isolates.	AN محما	NALYSIS						
	¢ ci	JSTOMISE	-					
	Locus Schem	al options display e display e field display						
	<b>9</b> IN	FORMATION	+					
		PING						

Either select the locus id by querying for it directly.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Query locus information	
Query locus information	Help 🕜 🚺
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 que Search criteria	Jery settings.
Display: 25 v records per page ()	
1 record returned. Click the hyperlink for detailed information. ——Customize—— Locus options	
Id       Idata type       locus type       allele id format       length varies       coding sequence       isolate display       main display       query field       analysis*       view         abc2       DNA       partial CDS       integer       433       allele only       Image: Coding sequence       allele only       Image: Coding sequence       coding sequence       sequence	

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 lo		
Query locus information	Help 🔀 🔮	
	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 🗸 = 🗸	+ i Order by: id vascending v	
	Display: 25 v records per page (i)	
	Action	
data type:		
locus type:	RESET SEARCH	
allele id format:		
length varies: 💉 🧃		
coding sequence:		
complete cds:		
match longest:		
pcr filter:		
probe filter:		
introns:		
isolate display:		
query field:		
query field: 💙 👔 analysis: 💙 👔		
submission template:		
view:		
curator:	<u> </u>	
scheme: MLST	× 0	
7 records returned. Click the hyperlinks for detailed information.		
Customize		
Locus options		
id  data type  locus type  allele id format length length varies c	coding sequence isolate display* main display* query field* analysis* view	
abcZ DNA partial CDS integer 433	allele only 🖸 🗹	
adk     DNA     partial CDS     integer     465       aroE     DNA     partial CDS     integer     490	Image: state of the state o	
fumC DNA partial CDS integer 465	allele only D V	
gdh         DNA         partial CDS         integer         501           pdhC         DNA         partial CDS         integer         480	Image: Second se	
pgm DNA partial CDS integer 450	Image: Second se	
* Default values are displayed for this field. These may be overridden by user	i preierence.	
Once loci have been selected, click Custor	mize locus options.	
scheme: MLST	<b>~</b> 0	
7 records returned. Click the hyperlinks for detailed information.		
Customize		
Locus options		
Locasophons		
id data type locus type allele id format length length varies codir	ling sequence isolate display* main display* query field* analysis* view	

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		$\checkmark$	$\checkmark$	
adk	DNA	partial CDS	integer	465			allele only		$\checkmark$	$\checkmark$	
aroE	DNA	partial CDS	integer	490			allele only		$\checkmark$	$\checkmark$	
fumC	DNA	partial CDS	integer	465		$\checkmark$	allele only		$\checkmark$	$\checkmark$	
gdh	DNA	partial CDS	integer	501		$\checkmark$	allele only		$\checkmark$	$\checkmark$	
pdhC	DNA	partial CDS	integer	480			allele only		$\checkmark$	$\checkmark$	
pgm	DNA	partial CDS	integer	450		$\checkmark$	allele only		$\checkmark$	$\checkmark$	
* Defau	lt values are	e displayed for	this field. These m	av be ov	erridden by use	r 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

## TWELVE

# **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 spe database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequent sample.			n						
Query database		DI LOG IN							
Search database									
Browse, search by any criteria, or enter list of attributes.		PROJECTS	-						
		Public projects Your projects							
Search by combinations of loci		Tour projects							
This can include partial matches to find related isolates.		EXPORT	+						
		MALYSIS							
			+						
			+						
		S TYPING							

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	14 <b>(</b>
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: UK group W ST-11 2016 Description: Group W ST-11 isolates from the UK in 2016. CREATE	
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 > Neisseria s	pp. > Neisseria isolates > Sear	ch or browse database					
Search or brows	se database					🛓 Help 🖍	
Enter search criteria or leave bla	ank to browse all records. Modify	form parameters to filter or e	nter a list of values.				
	5		Allele designations/so	heme fields			
Combine with: AND V		S	T (MLST)	=	✓ 11	+ ()	
country 🗸 :	- VK						
year 🗸 :	<b>~</b> 2016		Display/sort options -				
capsule group 🗸 :	- • W		Order by: id Display: 25 v re			✓ ascending ✓	
RESET SEARCH							
196 records returned (1 - 25 dis	played). Click the hyperlinks for	letailed information.					
— Your projects	Boc	kmark query					
UK group W ST-11 2016 ~	Add these records 2020-	07-29:1 Add b	ookmark				
(<) (<) 1 2 3 4	4 5 6 🕥 📎						
	Isolate fields (i			MLST	Finetyping and	igens	
id isolate aliases c	ountry year disea		capsule group		PorA_VR1 PorA_VR		
42451 M16 240077	UK 2016	Neisseria menin		1 ST-11 complex		F1-1	
	[England] 2016 invasive (unspe [England] 2016 invasive (unspe			1 ST-11 complex 1 ST-11 complex		F1-1 F1-1	
	[England] 2016 Invasive (unspe [England] 2016 invasive (unspe			1 ST-11 complex		F1-1 F1-1	
	[England] 2016 invasive (unspe			1 ST-11 complex		F1-146	

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
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     WK 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.

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	2	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 affects 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 42480 42483 42484 42483 42484 42485 42485 42485 42485 42485	; which		
You can also add isolate records to this project from the results of a query.			

# 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
Your projects
Delete     Add/remove records     Modify users     Project     Description     Administrator     Isolates     Browse       X     Image: Add/remove records     UK group W ST-11 2016     Group W ST-11 isolates from the UK in 2016.     Image: Add/remove records     Image: Add/remove remove records     Image: Add/remove records
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 bro	owse database				
Search or browse database				💄 Help 🕜 🚯	
Enter search criteria or leave blank to browse all records. Modify form pa	rameters to filte	r or enter a list of values.			
Isolate provenance fields		Filters			
id v = v Enter value	+ ()	Publication:	Select options	<b>`</b> ()	
		Project:	1 selected	1	
		× MLST profiles:			
		× Clonal complex (MLST):		© <u> </u>	
		Sequence bin:		^ <b>`</b> (i)	
		Include old record version Add filter:	2010/11-2013/14	✓ Add	
Display/sort options		Action	✓ UK group W ST-11 2016	Add	
	✓ ascending	RESET SEAF	UKMenCar2		
Display: 25 🗸 records per page 👔			UKMenCar3		
			7000 Zooniverse	<b>_</b> _	

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
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     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.

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	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	
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 T	
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: State of the state of	
Update users	
Add user	
Enter username: hbratcher You need to know the username of 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	Help 🔽
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/fremove records     Modify users     Project     Description     Administrator     Isolates     Browse       Image: State of the state	
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	3
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	
E 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.     Image: Complex of the user of t	
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

## THIRTEEN

# **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 > Organis	ms > Neisseria spp. > Neisseria isolates		
Neissei	<i>ia</i> isolates database		<u></u>
	Query database	€ LOG OUT	+
	Search database		
	Browse, search by any criteria, or enter list of attributes.	🔒 PRIVATE DATA	
		PROJECTS	+
	Search by combinations of loci This can include partial matches to find related isolates.	EXPORT	+
		MALYSIS	
			+
			+

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	ß
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 Currate 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 your que (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	
<ul> <li>Instructions</li> <li>This page allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.</li> <li>Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.</li> <li>Optionally enter references for your isolates as a semi-colon () separated list of PubMed ids (non-integer ids with 2000 and allele fields along with the other isolate data - simply create a new column with the locus in These will be added with a confirmed status and method set as 'manual'.</li> <li>Ou can choose whether or not to include an id number field - if it is omitted, the next available id will be used at Despetable if its isomitted.</li> <li>Example to the description of database fields for help with filling in the template.</li> </ul>	ame (see the 'allowed_loci' tab in the Excel template for locus names).
Upload Paste in tab-delimited text ( <b>include a field header line</b> ).	Action
	RESET

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

# 13.2 Modifying private records

Click the 'Private data' link on the contents page.

BIGSdb			
Home > Organis	ms > Neisseria spp. > Neisseria isolates		
Neissei	<i>ia</i> isolates database		ă l
	Query database	€ LOG OUT	+
	Search database		
	Browse, search by any criteria, or enter list of attributes.	A PRIVATE DATA	
		PROJECTS	+
	Search by combinations of loci This can include partial matches to find related isolates.	EXPORT	+
			+
		<b>1</b> INFORMATION	+

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.

BIGSd	b	
Home >	Organisms > Neisseria spp. > Neisseria isolates > Private records	
Priv	ate records	Help 🗹
Account	Limits	
,	Records (total): 0 Records (quota): 0 Quota: 100 You can upload: 100	
1	Upload 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 <b>not be private</b> . Either use the 'Upload' link above to upload private rec (if you have any available), or use the upload link on the individual projects listed below to upload to those projects.	ords to your 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

# FOURTEEN

# 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 > Organis	ms > 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	+
		C EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.	🗠 ANALYSIS	
		CUSTOMISE	+
			+
		S TYPING	

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.

	00/0/0/				intrastre (unspectied, other)		0		ST SE COMPLEX		10	
.3	139M	B99; NIBSC_2795; Z10	99 Philippines	1968		Neisseria meningitidis	А	1	ST-1 complex	5-2	10	F5-1
4	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitidis	Х	117			14	
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	174	ST-174 complex			
8	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
9	S3131	B213; NIBSC_2813; Z1	13 Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-
0	5		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
1	6		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
2	7		Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningitidis	В	32	ST-32 complex			
			27 Denmark	1074	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	5-1	9	F3-
	S4355	B227; NIBSC_2806; Z1	27 Definitian		invasive (unspecified/outer)							
	\$4355 9	1 2 3 4 5	Germany	1999	carrier	Neisseria meningitidis	В	930	ST-334 complex			
425		1 2 3 4 5					В	930	ST-334 complex			
5 C	, 3 (	1 2 3 4 5 tools	6 () (%)	1999			В	930	ST-334 complex			
25 (~ A1	9	1 2 3 4 5 tools	6 () (%)	1999	carrier	Neisseria meningitidis	В	930	ST-334 complex			

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.

Home > Organisms > Neisseria spp.	> Neisseria isolates > Plugins > BLAST	
BLAST		Help 🕜
Please select the required isolate ids	to BLAST against (use CTRL or SHIFT to make multiple selections) and paste in your query sequence. Nucleotide or peptide sequence	ences can be queried.
1 2 3 4 5 6 7 7 Clear List all	GAAGCCTTAGGCGGACGCGATGCCGCCTGCCGCTTCGGGCGCATTGAAAACGCTGGCGGCAAGCCTGAA TAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCCCGCGTTGCGGTTTGGGCGAAATCAAAATCCCCG AAAACGACCCGGGTTCGTCCATCATGCCGGCGAGTCCAACCCGACCCCAATGCGAAGCGAGCG	
Include in results table country region year date sampled isoyear sampled date received non culture epidemiological year age yr	Parameters         BLASTN word size:         BLASTN scoring:         reward.2; penalty:-3; gap open:5; gap extend:2 ▼ ①         Hits per isolate:         Hits per isolate:         Image: State isolate:         Image: State isolate:         BLASTN Scoring:         Restrict included sequences by         Options         Restrict included sequences by         Show isolates with no matches         Sequence method:         Include sequin id and start position in FASTA         Project:	)
Action	Experiment: 🔍 🗸 🕜	

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.

	Latab (	viewer										
	Status											
			db_009799_159 )-07-29 15:33:30		173							
		Status: finish	ned )-07-29 15:36:21									
		ogress: 100										
			0-07-29 15:36:29	)								
	Iota	al time: 7 sec	onds									
Outpu	ıt											
Isolate i	d Isolate	% identity A	lignment lengti	Mismatche	s Gaps	Seabin id	Start	End	Orientatio	n 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 6748	98.280 98.280	465 465	8	0	180965 181406	19869 36669	20333 extract → 37133 extract →	$\rightarrow$	0.0	803 803	
10	129E	98.280	465	8	0	181408	35889	36353 extract →	$\rightarrow$ $\rightarrow$	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		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 98.280	465 465	8	0	182815 182880	19783 18879	20247 extract → 19343 extract →	$\rightarrow$ $\rightarrow$	0.0	803 803	
52	243	98.280	465	8	0	182880	4532	4996 extract →	→ ←	0.0	803	
61	393	98.280	465	8	0	183648	7217	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 90	IAL2229 CN100	98.280 98.280	465 465	8	0	184635 184814	4530 5350	4994 extract → 5814 extract →	→ ↓	0.0	803 803	
Files	01100	70.200	105	0	U	101014	5550	Joz + extract -	<i>—</i>	0.0		
1 nes												
	FASTA				FAST	A with flar	nking		L Tab	le (tab-de	imited text)	
X				X					∎ Tab	(	,	
	Table (E)	xcel format)				ile containi	ing all outp	out				
x					files							

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

Output											
1							-				
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.

BI	IGSdb
Но	me > Organisms > Neisseria spp. > Neisseria isolates > Extracted sequence
E	Extracted sequence: Seqbin id#:180965 (19869-20333)
:	Sequence position       Options         sequence bin id:       180965         sequence method:       Solexa         start:       19869         end:       20333         length:       465         orientation:       →
	Sequence CGGCACGGGT TIGAACAGCC ATCCCGAATA CGCCGAAAAA GCCGCCGCA AACTCGCCGA ATTGTCCGGC TIGCCGTTTG TCAGCGCGCC GAACAAATTT GAAGCCTTGG GCGGGACGCGA TGCCGCCGTT GCCGCTTCGG GCGCATTGAA AACGCTGGCG GCAAGCCTGA ATAAAATCGC CAACGACATC CGCTGGCTGG CAAGCGGCCC GCGCTGCGGT TTGGGGCGAAA TCAAAATCCC CGAAAACGAG CCGGGTTCGT CCATCATGCC GGGCAAAGTC AACCCGACCC AATGCGAAGC GATGACCATG GTGTGCTGC AAGTGTTCGG CAACGACGTT ACCATCGGTA GTCGGGCAAT TICGAGCTGA ACGTCTATAT GCCCGTCATC GCTACAACC TCTTGCCGTT TGGGCGACG GTGGGCGCG CGTGCAACAG CTTCAACGAA ACCGGCCGT CGGGCAATTGA ACCCCGTACCG GAAAAAATCG ACTATTTCCT GCACCATTCC CTGATGCTGG TTACCGGTTAAACGCCAAA ATCGGTTACG AAAAACGCCCC AAAACCGCC ACAAAAAACGACAA ACCCGTACCG GAAAAAATCG CCGTTGAGTT GGGCTTGCTG ACGGGCGAAG AATTTGACGA ACTGG Translation
	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 FI 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 I CGGCACGGGTTTGAACAGCCATTCCCGAATAGCCGCCGCAATGTCGCGCGATGTCGCGGTTGCCGGCTTGCCGGCGAACAAGTT 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.

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 →	$\leftarrow$	0.0	803	
Files	FASTA				FAST	Ă with flan	king		E Tabl	e (tab-del	limited text	
X	Table (Ex	cel format)			Tar fi files	le containii	ng all out	put				
Please not	e that job re	esults will ren	nain on the serv	ver 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.

Include in results table	
country	A
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.

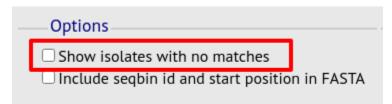
## 14.1.2 Parameters fieldset

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.

	ences by	
Sequence method:	<b>~</b> (i)	
Project:		► (i)

Available options are:

- Sequence method Choose to only analyse contigs that have been generated using a particular method. This depends on the method being set when the contigs were uploaded.
- Project Only include isolates belonging to the chosen project. This enables you to select all isolates and filter to a project.

# **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

	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningitidi	is B	1015	ST-32 complex	7	16	
13	139M	B99; NIBSC_2795; Z1099	Philippines	1968	,	Neisseria meningitidi		1	ST-1 complex	5-2	10	F5-1
14	0120/95	_	Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitidi	is X	117			14	
15	1		Germany	1999	carrier	Neisseria meningitidi	is E	864				
16	2		Germany	1999	carrier	Neisseria meningitidi	is B	854	ST-18 complex			
17	3		Germany	1999	carrier	Neisseria meningitidi	is W	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningitidi	is B	19	ST-18 complex			
19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidi	is A	4	ST-4 complex	7	13-1	F1-5
20	5		Germany	1999	carrier	Neisseria meningitidi			ST-198 complex			
21	6		Germany	1999	carrier	Neisseria meningitidi	is NG	198	ST-198 complex			
22	7		Germany	1999	carrier	Neisseria meningitidi	is E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningitidi		32	ST-32 complex			
24	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidi	is A	5	ST-5 complex	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria meningitidi	is B	930	ST-334 complex			
25	,					<u> </u>						
«	) 🕢	<b>1 2 3 4 5 6</b>	00	)								
(« Ar	)()	tools	Combinations	)	morphic sites Publication	s Sequence bin						
(« Ar	alysis f	tools	00	Poly	morphic sites Publication		PCR					

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	
This is the original BURST algorithm, developed by Ed Feil, first implemented by Man-Suen Chan. This version has been adapted for use as a plugin for the BIGSdb database software by Keith Jolley.	
BURST analysis can be used to:	
<ul> <li>Divide strains into groups according to their allelic profiles.</li> <li>Count the number of Single Locus Variants (SLV), Double Locus Variants (DLV) and Satellites (SAT) for each sequence type (ST).</li> <li>Identify the potential Ancestral Type (AT). These are shown with an asterisk next to their names in the results table.</li> </ul>	
Graphic representations of BURST groups can be saved in SVG format. This is a vector image format that can be manipulated and scaled in drawing packages, including the freely available Inkscape.	
Options Action	
Select scheme: MLST v 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	

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.

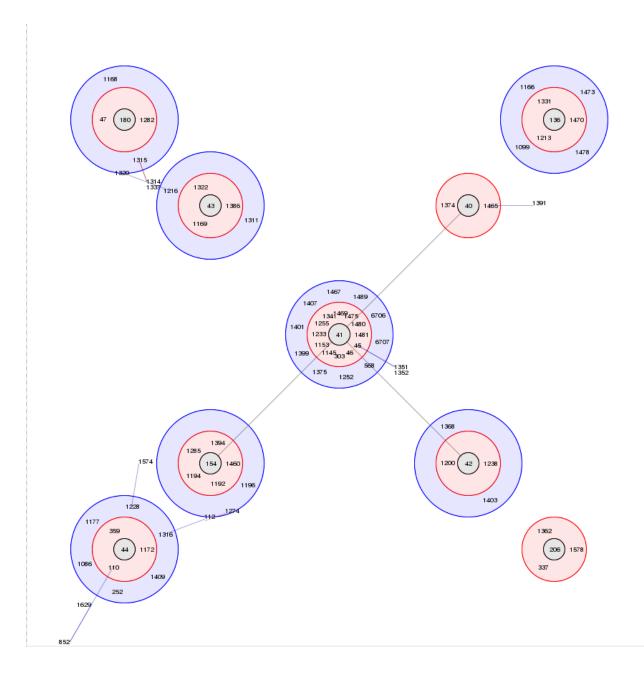
Click Submit.

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	•		
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
	484	1100		
	230 SVG file (right c	114	/	

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	DILOG IN	
Search database		
Browse, search by any criteria, or enter list of attributes.	PROJECTS	+
	Z EXPORT	+
Search by combinations of loci	🗠 ANALYSIS	
Search by combinations of loci This can include partial matches to find related isolates.	<ul><li>La ANALYSIS</li><li>✿ CUSTOMISE</li></ul>	+
		+++

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
11111/313
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 bigsdbreadthedocs.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.

								-							
15	1			Germany	1999	carri		Neisseria meningi		E	864				
16	2			Germany	1999	carri	ier	Neisseria meningit		В	854	ST-18 complex			
17	3			Germany	1999	carri	ier	Neisseria meningi		W	174	ST-174 complex			
18	4			Germany	1999	carri	ier	Neisseria meningi	tidis	В	19	ST-18 complex			
.9	S3131	B213; NIBSC_2	2813; Z1213	Ghana		invasive (unspe	ecified/other)	Neisseria meningit	tidis	Α	4	ST-4 complex	7	13-1	F1-5
20	5			Germany	1999	carri	ier	Neisseria meningi		NG	198	ST-198 complex			
1	6			Germany	1999	carri	ier	Neisseria meningi	tidis	NG	198	ST-198 complex			
22	7			Germany	1999	carri	ier	Neisseria meningi	tidis	E	60	ST-60 complex			
3	8			Germany	1999	carri		Neisseria meningit		В	32	ST-32 complex			
24	S4355	B227; NIBSC_2	2806; Z1227	Denmark	1974 i	invasive (unspe	ecified/other)	Neisseria meningit	tidis	A	5	ST-5 complex	5-1	9	F3-1
										В	930	ST-334 complex			
25	9			Germany	1999	carri	ier	Neisseria meningit	tials	Б	950	51-554 complex			
« An	) 🕢		4		1999	carri	ler	Neisseria meningi	tidis	D	930	51-554 complex			
« An	)()		4	Combinations		norphic sites	Publication	_	tidis	D	930	SI-SS4 complex			
	) 🕢	own: Fields	4 Codons	00	Polyn		Publications	_		B	730	SI-SS4 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 iso	lates > Plugins > Codon usage	
Codon usage analysis		Help 🔽 🚺
	ication of codons can only be achieved	ly loci that have a corresponding database containing sequences, or with sequences tagged, can be d for loci for which the correct ORF has been set (if they are not in reading frame 1). Partial sequnces nclude. Output is limited to 500 records.
Select ids	Loci	Schemes
Paste in list of ids to include, start a new line for each. Leave blank to include all ids.	165_FDNA 165_FRNA (SSU_FRNA) 235_FRNA abcZ abcZ (KEIS1015) aceF (NEIS1279) 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 han © Use sequences tagged from the bin O Use allele sequence retrieved from external (	dled: 👔 latabase	Select codon order: alphabetical C or G ending codons first
Do not include sequences with problem flage	ed (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 status viewer	
Job status viewer	•
Status           Job id:         BIGSdb_004053_1596101846_94417           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	
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) Tar file containing all output files	
Please note that job results will remain on the server for 7 days.	

# 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.		+
		EXPORT	+
	Search by combinations of loci	MALYSIS	
	This can include partial matches to find related isolates.	CUSTOMISE	+
			+
		E TYPING	

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		Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973 inv	vasive (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 inv	vasive (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			
<u> </u>		1 2 7 4 5 6				_						
«	) () nalysis (	1 2 3 4 5 6 tools	00	)								
(« An	) 🕢				ations Sequence bin							

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.

 Export:
 Dataset
 Contigs
 Sequences

 Third party:
 GrapeTree
 PhyloViz

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Field breakdown	
Field breakdown of dataset	Help 🔽
Isolate records: 66,602	Controls
Field selection	<b>(</b> •0
Select field: Country v List: ® fields O loci O schemes	Theme: Th
3,000 2,000 1,000 0	

Other fields can be chosen by selecting them in the dropdown list box.

Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Field breakdown	
Field breakdown of dataset	Help 📝
Isolate records: 66,602  Field select field country List:   Relation  Select field country  List:  Relation  Select field country  List:  Relation  Select field country  List:  Relation  Select field country  Select fiel	s t Natural Earth

You can also breakdown loci and schemes by clicking the appropriate button. This will re-populate the dropdown list.

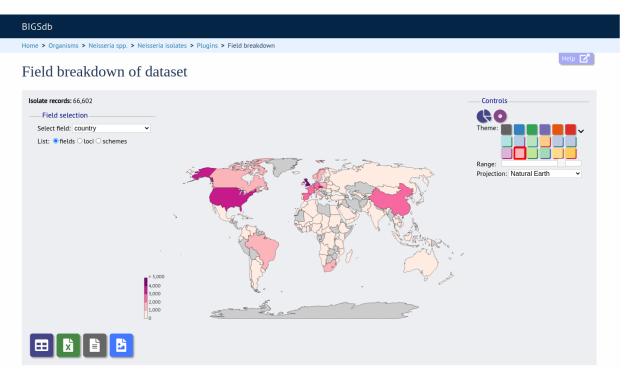
BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Field breakdown	
Field breakdown of dataset	Help 🔽
Isolate records: 66,602 Field selection Select field: clonal_complex (MLST)	Controls Max segments: 20
clonal_complex (55 values)	
11.2% 26.6% 25.5% 25.5% 25.5% 25.5%	
No value ST-11 complex ST-41/44 complex ST-32 complex ST-23 complex ST-23 complex ST-243 complex ST-213 complex ST-22 complex ST-32 complex ST	

The charts are dynamic and you can manipulate some aspects of them using controls shown on the screen.

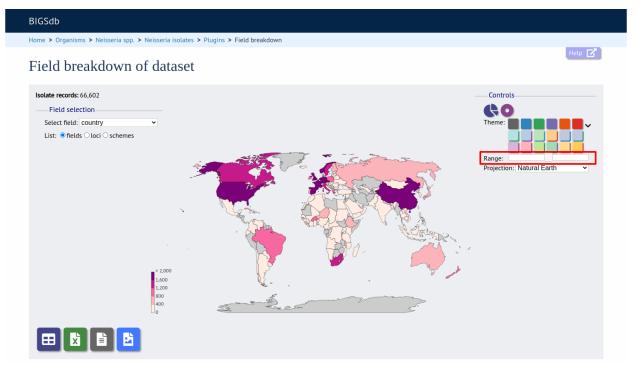
## 14.4.1 Maps

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.



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'.

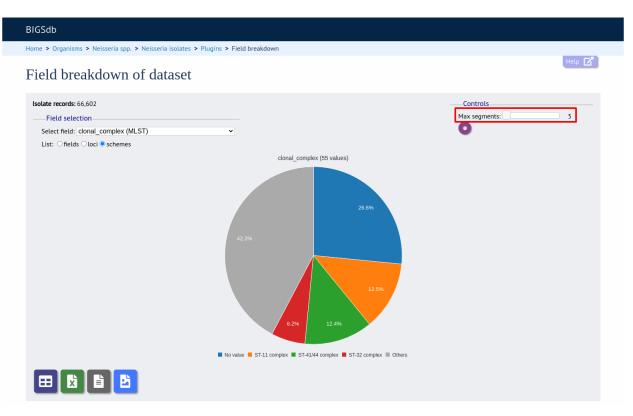


To see the same data as a pie chart, click the 'Pie' or 'Donut' icons.



## 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

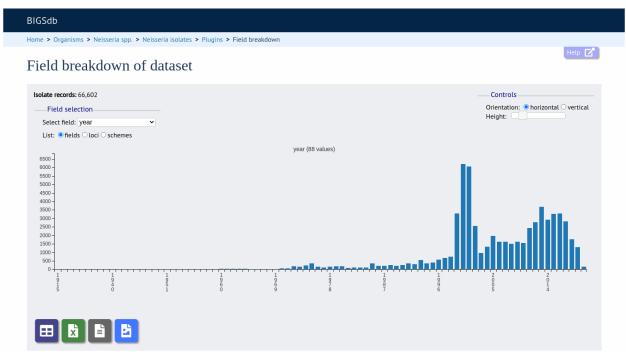
The icon changes to a pie chart image (clicking this will return to the pie chart).



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.

BIGSdb						
Home > Organisms > Neisseria spp. > Neisseria isolat	es > Plugins > Field breakdow	'n				
Field breakdown of datase	et					Help 🔽
Isolate records: 66,602						Controls
Field selection						Height:
Select field: date_entered						Lui
List: ● fields ○ loci ○ schemes						
70000 -		date_entered (2266 v	alues)			
60000 -						
50000 -						
40000 -						
30000 -						
20000 -						
10000 -						
0						
+++++++++++++++++++++++++++++++++++++++						
22005-06-05 	2007-08-04	2009-10-01	2011-11-30	2014-01-27	2016-03-27	
07 07 05	04	.01	30	.27	.27	26
الله الله الله						

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.

Sdb		
e > Organisms > Ne	isseria spp. >	Neisseria isolates
eakdown	by cou	ntry
1 values.		
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.

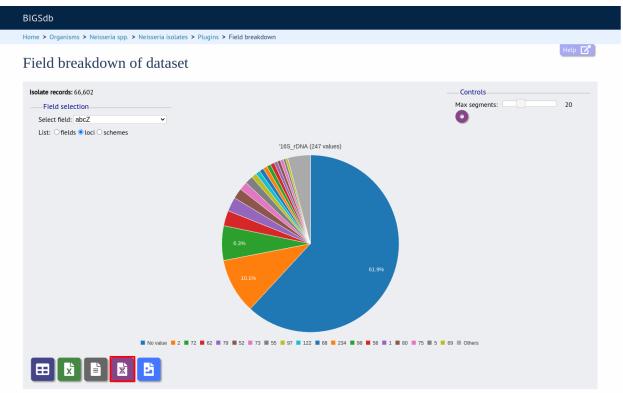


The chart image can also be saved as a SVG file, suitable for manipulation in a graphics program.



## 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		
I	Home > Organisms > Neisseria spp. > Neisseria isolates		
	<i>Neisseria</i> isolates database		
	Query database	→) 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.		
			+
			+

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 bigsdbreadthedocs.io 🗗
BICS/B Anne Algunes - Novem (g. ) - Novem Laters - Puger - Know preserve
Gene Presence - Neisseria isolates
Heitmap Congranditi Gonditi
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.

67         S5611         B466; NIBSC_2765; Z1466         Australia         1977         invasive (unspecified/other)         Neisseria meningitidis         A         1         ST-1 complex         5-2         10         F           82         11-004         B505; NIBSC_2786; Z1505         China         1984         invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           84         IAL229         B506; NIBSC_2786; Z1505         Brazil         1976         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           90         CN100         B534; NIBSC_2778; Z1515         Saudi Arabia         1987         carrier         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           20         F6498         NIBSC_2778; Z3515         Saudi Arabia         1987         carrier         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           20         F6498         NIBSC_2730; Z3524         Chad         1988         invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20 <t< th=""><th></th><th>393</th><th>B392; NIBSC_2823; Z1392</th><th>Greece</th><th>1968</th><th>carrier</th><th>Neisseria meningitid</th><th>lis A</th><th>1</th><th>ST-1 complex</th><th>5-2</th><th>10</th><th>F5-1</th></t<>		393	B392; NIBSC_2823; Z1392	Greece	1968	carrier	Neisseria meningitid	lis A	1	ST-1 complex	5-2	10	F5-1
32       11-004       B503; NIBSC_2826; Z1503       China       1984 invasive (unspecified/other)       Neisseria meningitidis       A       5       ST-5 complex       20       9       F         34       IAL2229       B506; NIBSC_2826; Z1503       Eval       1976       Neisseria meningitidis       A       5       ST-5 complex       20       9       F         34       IAL2229       B506; NIBSC_2816; Z1506       Brazil       1976       Neisseria meningitidis       A       5       ST-5 complex       20       9       F         34       IAL2229       B506; NIBSC_2731; Z5514       UK [England] 1941       Invasive (unspecified/other)       Neisseria meningitidis       A       5       ST-5 complex       20       9       F         32       F6124       NIBSC_2731; Z5515       Saudi Arabia       1987       carrier       Neisseria meningitidis       A       5       ST-5 complex       20       9       F         32       F6124       NIBSC_2732; Z5524       Chad       1988 invasive (unspecified/other)       Neisseria meningitidis       A       5       ST-5 complex       20       9       F         10       H1964       NIBSC_27371       UK       1987       invasive (unspecified/other)       Neisseria	54	254	B439; NIBSC_2812; Z1439	Djibouti	1966	invasive (unspecified/other)	Neisseria meningitid	dis A	1	ST-1 complex	5-2	10	F1-7
84         IA.22229         BS06; NIBSC_2816; 21506         Brazil         1976         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           90         CN100         B534; NIBSC_2756; 21534         UK [England]         1941         invasive (unspecified/other) Neisseria meningitidis         A         2         ST-5 complex         20         9         F           120         F4698         NIBSC_2731; 25515         Saudi Arabia         1987         carrier         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           128         F6124         NIBSC_2730; 25324         Chad         1988 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F           120         F4698         NIBSC_2730; 25374         Chad         1988 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F           120         H1964         NIBSC_2736; 23771         UK         1987 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F           121         H1964         NIBSC_2733;	67	S5611	B466; NIBSC_2765; Z1466	Australia	1977	invasive (unspecified/other)	Neisseria meningitid	dis A	1	ST-1 complex	5-2	10	F5-1
90         CN100         B534; NIBSC_2768; Z1534         UK [England]         1941         invasive (unspecified/other)         Neisseria meningitidis         A         21         5-2         10         Ff           120         F4698         NIBSC_2731; Z5515         Saudi Arabia         1987         carrier         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           128         F6124         NIBSC_2731; Z5515         Saudi Arabia         1987         carrier         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           128         F6124         NIBSC_2730; Z5524         Chad         1988 invasive (unspecified/other) Neisseria meningitidis         A         4         ST-5 complex         7         13-1         F           120         H1964         NIBSC_2736; Z5771         UK         1987 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F           120         H1964         NIBSC_2736; Z5771         UK         1987 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F           123         T44/76         <	82	11-004	B503; NIBSC_2826; Z1503	China	1984	invasive (unspecified/other)	Neisseria meningitid	lis A	5	ST-5 complex	20	9	F3-8
L20         F4698         NIBSC_2731; Z3515         Saudi Arabia         1987         carrier         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           L28         F6124         NIBSC_2730; Z3524         Chad         1988 invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           L20         F4698         NIBSC_2730; Z3524         Chad         1988 invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           L20         H1964         NIBSC_2730; Z3571         UK         1987 invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           L20         H1964         NIBSC_2736; Z3711         UK         1987 invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           L237         H44/76         44/76-5; NIBSC_2724; Z3842         Norway         1976 invasive (unspecified/other)         Neisseria meningitidis         B         32         ST-32 complex         7         16         F <tr< td=""><td>84</td><td>IAL2229</td><td>B506; NIBSC_2816; Z1506</td><td>Brazil</td><td>1976</td><td></td><td>Neisseria meningitid</td><td>lis A</td><td>5</td><td>ST-5 complex</td><td>20</td><td>9</td><td>F2-1</td></tr<>	84	IAL2229	B506; NIBSC_2816; Z1506	Brazil	1976		Neisseria meningitid	lis A	5	ST-5 complex	20	9	F2-1
L28         F6124         NIBSC_2730; Z3524         Chad         1988 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F6           L60         1014         NIBSC_2821; Z3667         Sudan         1985 invasive (unspecified/other) Neisseria meningitidis         A         4         ST-4 complex         7         13-1         F6           L10         H1964         NIBSC_2796; Z3771         UK         1987 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F6           Z10         H1964         NIBSC_2796; Z3771         UK         1987 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F6           Z37         H44/76         44/76-53; NIBSC_2724; Z3842         Norway         1976 invasive (unspecified/other) Neisseria meningitidis         B         32         ST-5 complex         20         9         F6           Z38         153         NIBSC_2733; Z3905         China         1966 invasive (unspecified/other) Neisseria meningitidis         A         5         ST-5 complex         20         9         F6	90	CN100	B534; NIBSC_2768; Z1534	UK [England]	1941	invasive (unspecified/other)	Neisseria meningitid	dis A	21		5-2	10	F3-9
160         1014         NIBSC_2821; Z3667         Sudan         1985 invasive (unspecified/other)         Neisseria meningitidis         A         4         ST-4 complex         7         13-1         F           101         H1964         NIBSC_2796; Z371         UK         1987 invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         7         13-1         F           237         H44/76         44/76-53; INESC_2796; Z371         UK         1987 invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           238         153         NIBSC_2733; Z3905         China         1966 invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20         9         F           238         153         NIBSC_2733; Z3905         China         1966 invasive (unspecified/other)         Neisseria meningitidis         A         5         ST-5 complex         20         9         F	120	F4698	NIBSC_2731; Z3515	Saudi Arabia	1987	carrier	Neisseria meningitid	dis A	5	ST-5 complex	20	9	F3-1
VIDESC_2796; Z3771         UK         1987 invasive (unspecified/other)         Nelsseria meningitidis         A         5         ST-5 complex         20         9         F           237         H44/76         44/76-5; NIBSC_2724; Z3842         Norway         1976 invasive (unspecified/other)         Nelsseria meningitidis         B         32         ST-32 complex         7         16         F           238         153         NIBSC_2733; Z3905         China         1966 invasive (unspecified/other)         Nelsseria meningitidis         A         5         ST-5 complex         20         9         F	128	F6124	NIBSC_2730; Z3524	Chad	1988	invasive (unspecified/other)	Neisseria meningitid	lis 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       B       32       ST-32 complex       7       16       F         238       153       NIBSC 2733; Z3905       China       1966 invasive (unspecified/other) Neisseria meningitidis       A       5       ST-5 complex       20       9       F	160	1014	NIBSC_2821; Z3667	Sudan	1985	invasive (unspecified/other)	Neisseria meningitid	lis A	4	ST-4 complex	7	13-1	F1-5
238 153 NIBSC_2733; Z3905 China 1966 invasive (unspecified/other) Neisseria meningitidis A 5 ST-5 complex 20 9 F	210	H1964	NIBSC_2796; Z3771	UK	1987	invasive (unspecified/other)	Neisseria meningitid	dis A	5	ST-5 complex	20	9	F3-1
	237	H44/76	44/76-3; NIBSC_2724; Z3842	Norway	1976	invasive (unspecified/other)	Neisseria meningitid	dis B	32	ST-32 complex	7	16	F3-3
	238	153	NIBSC_2733; Z3905	China	1966	invasive (unspecified/other)	Neisseria meningitid	lis A	5	ST-5 complex	20	9	F3-1
	«	)()	1 2 3 4 🕥	<b>&gt;&gt;</b>									
Analysis tools	« Ar	) 🔇		<b>&gt;&gt;</b>									

Sreakdown:	Fields	Two Field	Combinatio	ons Polym	orphic sites	Publication	s Sequence bin	
🛃 Analysis:	BURST	Codons	Gene Presen	ce Genom	e Comparato	r BLAST	rMLST species id	PCR
Export:	Dataset	Contigs	Sequences					
🗹 Third party:	GrapeTre	e iTOL	PhyloViz	Microreact				

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.

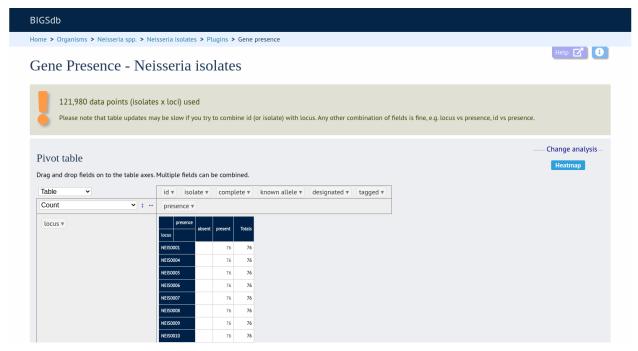
BIGSdb				
Home > Organisms > Neisseria spp. > Neisseria is	solates > Plugins > Gene preser	nce		
Gene Presence - Neisser Please select the required isolate ids and loci for loci defined in schemes by selecting the appropria Interactive analysis is limited to 500,000 data poi	comparison - use CTRL or SHIFT t ate scheme description.		s. In addition to selecting individual loci, you can choose to in	ielp 👔 🚺
Isolates User ger	nomes	Loci	Recommended schemes	
2 database. 7 Upload asse 10 Upload asse 11 Crip file of 13 FASTA files	nclude data not in the mbly FASTA file ontaining multiple - one per genome): 3 e No file chosen	165_rDNA 165_rRNA (SSU_rRNA) 235_rRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) 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	
Schemes	Parameters / options			
Alt loci     Genetic Information Processin     Genetic Information Processin     Genetic Information Processin     Genetic Islands     Geneti	Min % identity: 70 Min % alignment: 50 BLASTN word size: 20	) ~ ()		
Action RESET SUBMIT				

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         BiGSdb_017561_1596120664_37507           Submit time         2020-07-30 14:51:04           Status:         finished           Status:         100%           Status:         100%           Stop 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								
Home > Organisms > N	eisseria spp. > Neis	sseria isolates >	Plugins > Gen	e presence				
Gene Prese	nce - Neis	sseria is	olates					Help 🗹 🚺
	a points (isolates at table updates may		ry to combine i	d (or isolate) with lo	cus. Any other con	ibination of	fields is fine, e.g. locus vs presence, id vs presenc	e.
Pivot table Drag and drop fields of	n to the table axes. I	Multiple fields ca						—— Change analysis — Heatmap
Table ~ Count		id v isolate	v complete v	v known allele v	designated 🔻	tagged 🔻		
	× : ↔	presence v					_	
locus 🔻	• :	presence       presence       locus       NEIS0001	76 7	16				
locus v		presence abse	76 7	16 16				
locus v	v t	presence         abse           locus         NEISO001           NEISO004	76 7 76 7 76 7	16 16				
locus v	v t	presence       locus     abse       NEISO001     NEISO004	76 7 76 7 76 7 76 7 76 7	6 6				
locus v	• • •	presence         abse           locus         abse           NEISO001            NEIS000√            NEIS0005	76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7	6 6 6 6 6				
locus v	• • •	presence         abse           Locus         abse           NEISO001            NEIS0004            NEIS0006            NEIS0007	76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7           76         7	16 16 16				

The table will be re-drawn including these fields.

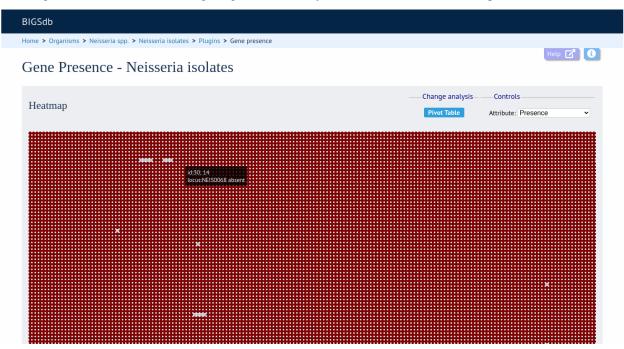
IGSdb											
ome > Organisms > Neisseria spp. > Ne	isseri	a isolate	s > P	lugins >	Gene p	resence					
Gene Presence - Nei	SS	eria	isc	lates	S						Help 🕜 🚺
121,980 data points (isolate Please note that table updates m				to combi	ne id (o	r isolate	) with	locus. An	iy other	combination of fields is fine, e.g. locus vs presence, id vs presence.	
Pivot table Drag and drop fields on to the table axes. Table	id		ate 🔻	be combi comple signated	ete 🔻	known a	allele	7			ange analysis Heatmap
locus 🔻	presence			absent		pres				1	
		designate	d not	designated	desi	Ignated	not d	esignated	Totals		
		tagged	— tagge	d untagged	tagged	untagged	tagged	untagged	Iotab		
	locu:	0001			76				76		
	-	0004			67		9		76		
	NEIS	0005			71		5		76		
	NEIS	0006			76				76		
	NEIS	0007			76				76		
	NEIS	0008			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 iso	lates > Plugins > Gene presence		
Gene Presence - Neisser	ia isolates		Help 🗹 🚺
Heatmap		Change analysis Pivot Table	Attribute: Complete sequences V
	1622(1) 53353 Incounterson	if 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.	PROJECTS	+
		EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.	🗠 ANALYSIS	
			+
			+
		S TYPING	

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

	Comparator		
mmary:	Compare genomes at	t defined loci or against loci defined in a reference genome	
mparato alysis fa at are.	r. Output is equivalen cilitates the determir	compare whole genome data of isolates within the database using either the database defined loci or th t to a whole genome MLST profile, a distance matrix calculated based on allelic differences and a Neigh nation of precisely which loci vary among isolates. Additionally, user-uploaded genomes that are not con	hborNet graph generated from this distance matrix. The
	ation bigsdb.readthedoo	sio 🗗	
Loci with sequence d	ifferences among holates		
Social         Public NM           HESSOD         Field           HESSODH         Field           HESSONH         Field           HESSONH         Field           HESSONH         Field           HESSONH         Field           HESSONH         Field	Anticipation of provide an anticipation and a second and	Base         Base <th< td=""><td></td></th<>	
(prE) NEIS0055 ctrA	capsele polysaccharide export outer membrane potain capsele polysaccharide export invermembrane exemption		

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.

	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitidis	Х	117			14	
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	174	ST-174 complex			
8	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
9	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex	7	13-1	F1-5
0	5		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
1	6		Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex			
2	7		Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
3	8		Germany	1999	carrier	Neisseria meningitidis	В	32	ST-32 complex			
4	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (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			

Breakdown: Fields Two Field Combinations Polymorphic sites Publications Sequence bin

### 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	Neisseria isolates > Plugins > Genome comp	parator
Genome Compar	ator - Neisseria isolates	Help 🗹 🚺
	rtefactual analysis because this database contains ised to select one of the recommended schemes.	s different kinds of loci that are not equivalent (complete genes, MLST fragments, peptide variable regions,
		make multiple selections in list boxes. In addition to selecting individual loci, you can choose to include all ly, you can enter the accession number for an annotated reference genome and compare using the loci
Isolates	User genomes	Loci
644 662 665 666 667	Optionally include data not in the database. Upload assembly FASTA file (or zip file containing multiple FASTA files - one per genome):	16S_rDNA 16S_rRNA (SSU_rRNA) 23S_rRNA abcZ abcZ abcZ (NEIS1015)
669 670	Choose file No file chosen	ALL None Paste list
Clear List all	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	schemes list. - I I Lineage Schemes - I I Metabolism - I I N. gonorrhoeae AMR - I I Plin
Annotated reference genom Enter accession number: or choose annotated genome: or upload Genbank/EMBL/FASTA Choose file No file chosen	Min % identity: 70 Min % alignment: 50 BLASTN word size: 20	Completely exclude from analysis     Include ref sequences in alignment       Treat as distinct allele     Align all loci (not only variable)       Create alignment states     Create alignment states
Core genome analysis	Filter by	Action
Core threshold (%): 90 💌 👔		

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.

ob stat	us view	ver											
Sta	itus												
- Otu		BIGSdb_02456	6 150619645	0 13471									
		2020-07-31 09		9_134/1									
		finished											
	Start time: Progress:	2020-07-31 09	:07:42										
	Stop time:	2020-07-31 09	:07:46		-								
	Total time:	4 seconds											
Output													
Output													
Analysis ag	gainst defined	lloci											
All loci													
Allele number	rs are used wher	e these have he	en defined of	henwise sea	uences will be	marked as 'New	v#1 'Now#7' o	to Missing alle	les are marker	las Y loco	molete alleler	s (located at e	ad of contia
are marked as	s T.	e these have be	en denneu, or	anerwise seq	dences witt be	marked as twee	v#1, Nevv#2 e	te. Phissing atte	tes are market	as A . Inco	inplete attetes	s (tocated at e	ia or contrag
	ame Product 64	4 (L93/4286) 6	62 (2837) 66	5 (2845) 66	6 (2843) 667 (	2842) 669 (284	6) 670 (2840	671 (2844) 6	572 (2847) 698	(FAM18) 417	84 (2839) 41	1785 (2838)	
abcZ adk												2 3	
aroE												4	
fumC gdh				8	8 8	8 8						8	
pdhC pgm												4 6	
	1:00												
LOCI WITH S	equence diffe	erences amon	ig isolates										
Variable loci: 4	4												
Locus Full na abcZ	ame Product 64	4 (L93/4286) 6	62 (2837) 66	5 (2845) 66	6 (2843) 667 (	2842) 669 (284	6) 670 (2840	671 (2844) 6	698	(FAM18) 417	84 (2839) 41	1785 (2838)	
aroE						4 4						4	
fumC						3 <u>3</u> 64	4	4	4	3 4	<u> </u>	<u> </u>	
pdhC		7											
pdhC Exactly ma	tching loci	-											
Exactly ma	_												
Exactly ma These loci are	tching loci e identical in all	isolates.											
Exactly ma These loci are Matches: 3	e identical in all												
Exactly ma These loci are Matches: 3	_		62 (2837) 66 3	5 (2845) 66 3	6 (2843) 667 ( 3	2842) 669 (284 5 3	6) 670 (2840 3	671 (2844) 6 3	572 (2847) 698 3	(FAM18) 417 3	84 (2839) 41 3	1785 (2838) 3	
Exactly mar These loci are Matches: 3 Locus Full na adk gdh	e identical in all	14 (L93/4286) 6 3 8	3 8		6 (2843) 667 ( 3 8	2842) 669 (284 5 3 8 8 6 6	-6) 670 (2840 3 6	671 (2844) 6 3 8	572 (2847) 698 3 6	(FAM18) 417 3 8 6	84 (2839) 41 3 6	1785 (2838) 5 6	
Exactly main These loci are Matches: 3 Locus Full na adk	e identical in all		62 (2837) 66 3 8 6	5 (2845) 661 3 8 6	6 (2843) 667 ( 3 3 8 4 6 4	2842) 669 (284 3 3 8 8 6 6	6) 670 (2840) 3 8 6	671 (2844) ( 3 8 6	572 (2847) 698 3 8 6	(FAM18) 417 3 8 6 9	84 (2839) 41 3 8 6	1785 (2838) 3 8 6	
Exactly mar These loci are Matches: 3 Locus Full na adk gdh	e identical in all	14 (L93/4286) 6 3 8	3 8		6 (2843) 667 ( 3 3 6 4	2842) 669 (284 3 3 8 8 6 6 6	6) 670 (2840) 3 8 6	671 (2844) 6 3 8 6	572 (2847) 698 3 8 6	(FAM18) 417 3 8 6	84 (2839) 41 3 8 6	1785 (2838) 3 8 6	
Exactly mai These loci are Matches: 3 Locus Full na adk gdh pgm	e identical in all	14 (L93/4286) 6 3 8	3 8		6 (2843) 667 ( 3 5 8 6	2842) 669 (284 3 8 8 6 6	6) 670 (2840) 3 8 6	671 (2844) 6 3 8 6	572 (2847) 698 3 8 6 9	6 (FAM18) 417 5 8 6	84 (2839) 41 3 8 6	1785 (2838) 3 8 6	
Exactly main These loci are Matches: 3 Locus Full na adk gdh pgm Unique strains Unique strains Strain 1	e identical in all ame Product 64 hins s: 5 Strain 2	14 (L93/4286) 6 3 6 6	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 5 6 6	2842) 669 (284 3 3 8 8 6 6	6) 670 (2840) 3 8 6	671 (2844) ( 3 8 6	572 (2847) 698 3 8 6	6 (FAM18) 417 3 8 6	84 (2839) 41 3 8 6	.785 (2838) 3 8 6	
Exactly mail These loci are Matches: 3 Locus Full na adk gdh pgm Unique strains Strain 1 662 (2837) 6	e identical in all ame Product 64 ains s: 5 Strain 2 544 (L93/4286) 1	14 (L93/4286) 6 3 8 6 5 5 5 5 6 6 6 6 6 5 6 6 9 5 5 5 6 9 5 5 5 6 9	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 8 4 6 6	2842) 669 (284 5 5 3 8 8 6 6 6	6) 670 (2840 3 8 6	671 (2844) ( 3 8 6	572 (2847) 698 5 8 6	6 (FAM18) 417 3 8 6	84 (2839) 41 3 8 6	.785 (2838) 3 8 6	
Exactly mar These loci are Matches: 3 Locus Full ns adk gdh pgm Unique strains Strain 1 662 (2837) 6 670 (2840) 671 (2844)	e identical in all ame Product 64 ains s: 5 Strain 2 648 (L93/4286) 1 698 (FAM18) 41784 (283)	14 (L93/4286) 6 3 8 6 5 5 5 5 6 6 6 6 6 5 6 6 9 5 5 5 6 9 5 5 5 6 9	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 3 6 6 0 0	2842) 669 (284 3 3 8 8 6 6 6	6) 670 (2840) 3 8 6	671 (2844) ( 5 8 6	572 (2847)   698 3	(FAM18) 417 3 8 6	84 (2839) 41 3 8 6	785 (2839) 3 8 6	
Exactly mar These loci are Matches: 3 Locus Full ns adk gdh pgm Unique strains Strain 1 662 (2837) 6 670 (2840) 671 (2844)	e identical in all ame <b>Product 6</b> 4 ams s: 5 <b>Strain 2</b> 698 (FAM18)	14 (L93/4286) 6 3 8 6 5 5 5 5 6 6 6 6 6 5 6 6 9 5 5 5 6 9 5 5 5 6 9	3 8 6 7	3 8 6 9	6 (2843) 667 ( 5 28 6 6	2842) (669 (284 3 3 8 8 8 6 6	6) 670 (2840) 3 8 6	671 (2844) ( 3 8 6	572 (2847)   695 3 8 6	(FAM18) 417 3 8 6	84 (2839) 41 3 8 6	785 (2838) 3 6	
Exactly mar These loci are Matches: 3 Locus Full ns adk gdh pgm Unique strains Strain 1 662 (2837) 6 670 (2840) 671 (2844)	e identical in all ame Product 64 ains s: 5 Strain 2 648 (L93/4286) 1 698 (FAM18) 41784 (283)	14 (L93/4286) 6 3 8 6 5 5 5 5 6 6 6 6 6 5 6 6 9 5 5 5 6 9 5 5 5 6 9	3 8 6 7	3 8 6 9	6 (2843) 667 ( 5 28 6 6	2842) 669 (284 5 3 8 8 5 6	6) 670 (2840 3 8 6	671 (2844) ( 3 8 6	572 (2847)) 698 5 6 6	(FAM18) 417 5 8 6	84 (2839)   41 5 8 6	<b>785 (2839)</b> 3 8 6	
Exactly ma These loci are Matches: 3 Locus Follin: adk gdh pgm Unique strains Strain 1 662 (2837) 6 670 (2840) 671 (2844) 672 (2847)	: identical in all ame Product 64 tins s: 5 Strain 2 44 (193/X486) 698 (FAM18) 41784 (2839) 41785 (2838)	14 (L93/4286) 6 3 8 6 5 5 5 5 6 6 6 6 6 5 6 6 9 5 5 5 6 9 5 5 5 6 9	3 8 6 7	3 8 6 9	6 (2843) 667 ( 5 2 6 0	2842) 669 (284 3 3 8 8 6 6	6) 670 (2840 3 8 6	671 (2844) ( 3 8 6	572 (2847)) 698 8 6	(FAM18) 417 5 8 6	84 (2839)   41 5 8 6	3 3 6	
Exactly ma These loci are Matches: 3 Locus Full ne adk gdh pgm Unique strains Strain 662 (2837) 6 662 (2847) 672 (2847) 672 (2847) 672 (2847) 672 (2847)	e identical in all ame <b>Product 64</b> hins s: 5 <b>Strain 2</b> 698 (FAM18) 41784 (2839) 41785 (2838)	41 (193/4226) ( 3 6 5 5 5 6 5 6 5 6 7 (2842) 5 6 7 (2842) 5 6 7 (2842) 5 6 7 (2842) 5 6 7 (2842) 5 6 7 8 8 8 8 8 9 8 8 8 8 8 8 8 8 8 8 8 8 8	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3	2842)   669 (284 5 3 8 8 6 6	6) 670 (2840 3 8 6	671 (2844) ( 3 8 6	572 (2847) (698 8 6	(FAM18) 417 3 8 6	84 (2839) 41 3 8 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Full ne adk gdh pgm Unique strain: Strain 1 662 (1837) 671 (2849) 671 (2849) 671 (2849) 671 (2849) Charts Click to enlarg	: identical in all ame Product 64 tins s: 5 Strain 2 44 (193/X486) 698 (FAM18) 41784 (2839) 41785 (2838)	41 (193/4226) ( 3 6 5 5 5 6 5 6 5 6 7 (2842) 5 6 7 (2842) 5 6 7 (2842) 5 6 7 (2842) 5 6 7 (2842) 5 6 7 8 8 8 8 8 9 8 8 8 8 8 8 8 8 8 8 8 8 8	3 8 6 7	3 8 6 9	6 (2843) (667 ( 3 3 3 3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	2842) 669 (284 5 5 8 8 6 6	6) 670 (2840 3 8 6	671 (2844) ( 3 8 6	572 (2847)   698 8 6	(FAM18) 417 3 8 6	84 (2839) 41 3 8 6	785 (2838) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Full ne adk gdh pgm Unique strain: Strain 1 662 (1837) 671 (2849) 671 (2849) 671 (2849) 671 (2849) Charts Click to enlarg	e identical in all ame <b>Product 64</b> hins s: 5 <b>Strain 2</b> 698 (FAM18) 41784 (2839) 41785 (2838)	41 (193/4226) ( 3 8 6 5 8 6 5 8 6 5 8 6 5 8 5 8 6 5 8 6 7 (2842) 5 8 5 8 6 7 (2842) 5 8 5 8 6 7 (2842) 5 8 6 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7 8	3 8 6 7	3 8 6 9	6 (2843)   667 ( 3 3 3 5 6 7 7 7 7 8 7 7 7 7 7 7 7 7 7 7 7 7 7 7	2842) 669 (284 5 5 8 8 6 6	6) 670 (2840 3 8 6	671 (2844) ( 3 6	572 (2847)   698 5 6	(FAM18) 417 3 8 6	84 (2839)   41 3 8 6	785 (2838) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Full ne adk gdh pgm Unique strain: Strain 1 662 (1837) 671 (2849) 671 (2849) 671 (2849) 671 (2849) Charts Click to enlarg	e identical in all ame <b>Product 64</b> hins s: 5 <b>Strain 2</b> 698 (FAM18) 41784 (2839) 41785 (2838)	41 (193/4226) ( 3 8 6 5 8 6 5 8 6 5 8 6 5 8 5 8 6 5 8 6 7 (2842) 5 8 5 8 6 7 (2842) 5 8 5 8 6 7 (2842) 5 8 6 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7 8	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 8 6	2842) 669 (284 5 5 5 8 8 6 6	6) 670 (2840 3 6	671 (2844) ( 3 6	572 (2847)   698 5 6	(FAM18) 417 3 8 6	84 (2839)   41 3 8 6	785 (2838) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Follone adk gdh ygm Unique strain: Strain 1 642 (2877) 6 670 (2840) 671 (2844) 671 (2844) 672 (2847) 6 Charts Click to enlarg	e identical in all ame <b>Product 64</b> hins s: 5 <b>Strain 2</b> 698 (FAM18) 41784 (2839) 41784 (2839) 998, 11785 (2838) ge.	41 (193/4226) ( 3 8 6 5 8 6 5 8 6 5 8 6 5 8 5 8 6 5 8 6 7 (2842) 5 8 5 8 6 7 (2842) 5 8 5 8 6 7 (2842) 5 8 6 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7 8	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 8 6	2842) 669 (284 5 3 8 8 6 6 6	6) 670 (2840 3 8 6	671 (2844) ( 3 8 6	572 (2847)   698 3 6 6	(FAM19) 417 3 8 6	84 (2839)   41 3 6 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Full ne adk gdh pgm Unique strain: Strain 1 662 (1837) 671 (2849) 671 (2849) 671 (2849) 671 (2849) Charts Click to enlarg	e identical in all ame   Product   64 ame   74 ame	4 (193/4286) (6 5 6 5 6 5 6 5 6 5 6 6 7 (2842) 9 4 9 4 4 1 1 1 1 1 1 1 1 1 1 1 1 1	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 8 6	2842) 669 (284 5 3 8 8 6 6 6	6) 670 (2840 3 6	671 (2844) ( 3 8 6	572 (2847)   698 3 6 6	(FAM19) 417 3 8 6	84 (2839)   41 3 6 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Follone adk adh pgm Unique strains Strain 5 662 (1837) 662 (1837) 662 (1837) 677 (1840) 677 (1840) 677 (1840) Charts Click to enlarge	e identical in all ame Product 64 ains s: 5 Strain 2 544 (193/4286) 698 (FAM18) 41784 (2839) 41785 (2838) ge. ocus presence fre	4 (193/4286) (6 5 6 5 6 5 6 5 6 5 6 6 7 (2842) 9 4 9 4 4 1 1 1 1 1 1 1 1 1 1 1 1 1	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 8 6	2842) 669 (284 5 3 8 8 6 6 6	6) 670 (2840 3 6	671 (2844) ( 3 8 6	572 (2847)   698 5 6	(FAM18) 417 3 8 6	84 (2839)   41 3 8 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Follone adk adh pgm Unique strains Strain 5 662 (1837) 662 (1837) 662 (1837) 677 (1840) 677 (1840) 677 (1840) Charts Click to enlarge	e identical in all ame   Product   64 ame   74 ame	4 (193/4286) (6 5 8 6 5 5 6 5 6 6 7 (2842) 9 4 (2842) 9 4 (2842) 9 1 1 1 1 1 1 1 1 1 1 1 1 1	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 8 6	2842) 669 (284 5 5 8 8 6 6 6	6) 670 (2840 3 8 6	671 (2844) ( 3 8 6	572 (2847)   698 3 8 6	(FAM19) 417 3 8 6	84 (2839) 41 3 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Full nei adk gdh pgm Unique strains Strain 4 670 (2840) 672 (2847) 672 (2847) 672 (2847) Charts Click to enlarge	e identical in all ame   Product   64 ame   74 ame	4 (193/4286) (6 5 8 6 5 5 6 5 6 6 7 (2842) 9 4 (2842) 9 4 (2842) 9 1 1 1 1 1 1 1 1 1 1 1 1 1	3 8 6 7	3 8 6 9	6 (2843) 667 ( 3 0 0 0 0 0	2842) 669 (284 5 5 8 8 6 6 6	6) 670 (2840) 3 8 6	671 (2844) ( 3 8 6	572 (2847)   698 3 6 6	(FAM19) 417 3 6 6	84 (2839) 41 3 8 6 2	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Full ne adk gdh pgm Unique strains Storial are 670 (2840) 672 (2847) 672 (2847) 672 (2847) 672 (2847) 672 (2847) 672 (2848) 672 (2840) 672	e identical in all ame   Product   64 ame   74 ame	4 (193/4286) (6 5 8 6 5 5 6 5 6 6 7 (2842) 9 4 (2842) 9 4 (2842) 9 1 1 1 1 1 1 1 1 1 1 1 1 1	3 8 6 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	5 8 6	6 (2843) 667 ( 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	5 3 8 8 6 6	6) 670 (2840) 3 8 6	3 6	Splits g	5 8 6	5 8 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Full nd adk gdh pgm Unique strains Strain 4 662 (2837) 6 670 (2840) 671 (2844) 672 (2847) 5 Click to enlarge Click to enlarge Locus Full nd adk gdh gdh gdh gdh gdh gdh gdh gdh	e identical in all ame Product 64 alins s: 5 Strain 2 544 (193/4786) 698 (FAM18) 41784 (2839) 41785 (2838) ge. presence free presence free presence in the diad	4 (193/4286) (6 5 8 6 5 5 6 5 6 6 7 (2842) 9 4 (2842) 9 4 (2842) 9 1 1 1 1 1 1 1 1 1 1 1 1 1	3 8 6 7	5 8 6	3 6 6	5 3 8 8 6 6	6) 670 (2840) 3 8 6	3 6	3 8 6	5 8 6	5 8 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Full ne adk gdh pgm Unique strains Storial are 670 (2840) 672 (2847) 672 (2847) 672 (2847) 672 (2847) 672 (2847) 672 (2848) 672 (2840) 672	e identical in all ame Product 64 alins s: 5 Strain 2 544 (193/4786) 698 (FAM18) 41784 (2839) 41785 (2838) ge. presence free presence free presence in the diad	4 (193/4286) (6 5 8 6 5 5 6 5 6 6 7 (2842) 9 4 (2842) 9 4 (2842) 9 1 1 1 1 1 1 1 1 1 1 1 1 1	3 8 6 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	5 8 6	3 6 6	5 3 8 8 6 6	6) 670 (2840) 3 8 6	3 6	Splits g	5 8 6	5 8 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Foll nei adk gdh pgm Unique strains Stora (2840) 670 (2840) 672 (2847) eff 672 (287)	e identical in all ame Product 66 Strain 2 Strain 2	4 (193/4236) 6 5 6 5 5 6 5 5 6 5 5 6 6 6 6 7 (2842) 6 6 6 7 (2842) 6 6 7 (2842) 6 6 7 7 8 6 7 8 8 8 7 8 8 8 8 8 8 8 8 8 8 8 8 8	3 8 6 (2846) 665 (2 (2846) 665 (2 )	5 8 6 10 2845) 2845) Excel	5 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	5 3 8 8 6 6 6	3 8 6	3 6	Splits g	5 8 6	5 8 6	785 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Foll nei adk gdh pgm Unique strains Stora (2840) 670 (2840) 672 (2847) eff 672 (287)	e identical in all ame Product 64 statistics string Strain 2 Strain 2	4 (193/4236) 6 5 6 5 5 6 5 5 6 5 5 6 6 6 6 7 (2842) 6 6 6 7 (2842) 6 6 7 (2842) 6 6 7 7 8 6 7 8 8 8 7 8 8 8 8 8 8 8 8 8 8 8 8 8	3 8 6 (2846) 665 (2 (2846) 665 (2 )	5 8 6 10 2845) 2845) Excel	5 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	5 3 8 8 6 6 6	3 8 6	3 6	Splits g	5 8 6	5 8 6	735 (2836) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Foll nei adk gdh pgm Unique strains Stora (2840) 670 (2840) 672 (2847) eff 672 (287)	e identical in all ame Product 66 Strain 2 Strain 2	4 (193/4236) 6 5 6 5 5 6 5 5 6 5 5 6 6 6 6 7 (2842) 6 6 6 7 (2842) 6 6 7 (2842) 6 6 7 7 8 6 7 8 8 8 7 8 8 8 8 8 8 8 8 8 8 8 8 8	3 8 6 (2846) 665 (2 (2846) 665 (2 )	5 8 6 10 2845) 2845) Excel	5 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	5 3 8 8 6 6 6	3 8 6	3 6	Splits g	5 8 6	5 8 6	795 (2839) 3 8 6	
Exactly ma These loci are Matches: 3 Locus Follor at adk gdh pgm Unique strains Strain 4 670 (2840) 672 (2847) at Charts Click to enlarge Click to enlarge Click to enlarge Files Files	e identical in all anne Product 64 strain 2 strain 2	4 (193/4236)         5           5         6           Strain 3         51           6         6         7           guency           0         0           0         0         0           guency           0         0           (Nexus format)           the number of	3 8 6 (2846) 665 (2 (2846) 665 (2 )	3     3       6     3       6     3       2345)     2345       coading in to erent allele s	5 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	s 3 8 8 6 6	3 8 6	3 6	Splits g Sylits (	5 8 6	5 8 6	795 (2839) 3 6	
Exactly ma These loci are Matches: 3 Locus Follor at adk gdh pgm Unique strains Strain 4 670 (2840) 672 (2847) at Charts Click to enlarge Click to enlarge Click to enlarge Files Files	e identical in all ame Product 64 s: 5 Stroin 2 544 (1937/4286) 698 (FAMI8) 41784 (2839) 41785 (2838) ge. ocus presence fre pesent in % of taok Present in % of taok	4 (193/4236)         5           5         6           Strain 3         51           6         6         7           guency           0         0           0         0         0           guency           0         0           (Nexus format)           the number of	3 8 6 (2846) 665 (2 (2846) 665 (2 )	5 8 6 10 2845) 2845) Excel toading in to	3 3 6 7 6 7 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	s 3 8 8 6 6	3 8 6	3 6	Splits g Sylits (	5 8 6	5 8 6	795 (2839) 3 6	

#### 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 comparator
Genome Comparator - Neisseria isolates
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.
Please select the required isolate ids and loci for comparison - use CTRL or SHIFT to make multiple selections in list boxes. In addition to selecting individual loci, you can choose to include all loci defined in schemes by selecting the appropriate scheme description. Alternatively, you can enter the accession number for an annotated reference genome and compare using the loci defined in that.
Isolates User genomes Include in identifiers
644     Optionally include data not in the database.     isolate country       665     Upload assembly FASTA file (or zip file containing multiple 669     year date sampled isoyear sampled       Clear List all     Choose file No file chosen
Annotated reference genome Parameters / options Distance matrix calculation Alignments
Enter accession number:       Min % identity: 70 v ()       With incomplete loci:       Produce alignments ()         or choose annotated genome:       Min % alignment: 50 v ()       Include ref sequences in alignment         FAM18 (Nm) v       BLASTN word size: 20 v ()       Include ref sequences in alignment         or upload Genbank/EMBL/FASTA file:       Rescan undesignated loci ()       Exclude pairwise missing loci         Choose file No file chosen       I
Core genome analysis Filter by Action
Core threshold (%): 90 v () Sequence method: () v () Project: () VERT SUBMIT Calculate mean distances () Project: () v () Experiment: v ()

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 referen																
	e genome															
access	ion: AM421808															
vers	ion: 1															
t	/pe: dna															
len	ath: 2194961															
descript		M18 complet	e denome													
coding regi		and complete	e genome.													
county regi	2,775															
ach unique allele is def	ined a number starting at 1. Missing alleles	are marked as	s <mark>'X'</mark> . Inco	omplete all	eles (locateo	l at end	of cont	ig) are n	arked a	as T						
ch 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 (2839)	
Locus		Sequence	Genome	Reference	644	662	665	666	667	669						
Locus MC0001 lpxC envA	Product UDP-3-O-[3-hydroxymyristoyl] N-	Sequence length	Genome position	Reference	644	662	665	666	667	669						
Locus	Product UDP-3-O-[3-hydroxymyristoyl] N- acetylglucosmine deacetylase	Sequence length 924	Genome position 1261	Reference genome 1	644 (L93/4286) 1	662 (2837) 1	665 (2845) 1	666 (2843) 1	667	669 (2846) 1	(2840) 1	(2844) 1	(2847) 1	(FAM18) 1	(2839) 1 2 2	
Locus MC0001 lpxC envA 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	644 (L93/4286) 1	662 (2837) 1	665 (2845) 1	666 (2843) 1 1 2 2	667 (2842) 1 1	669 (2846) 1 1	(2840) 1 1	(2844) 1 2 2	(2847) 1 1	(FAM18) 1	(2839) 1 2 2 2	
Locus MC0001 lpxClenvA NMC0002 pilS1 NMC0003 pilS2 NMC0004 fbp NMC0005	Product UDP-3-0-[3-hydroxymyristoyl] N- acetylglucosmine deacetylase pilin (fragment) truncated pilin peptidyl-prolyl cis-trans isomerase putative membrane protein	Sequence length 924 291 366 330 219	Genome position 1261 3341 3675 4069 4476	Reference genome 1 1 1	644 (L93/4286) 1 1 2	662 (2837) 1 2 2 3	665 (2845) 1 2 2 3	666 (2843) 1 2 2 3	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	(FAM18) 1	(2839) 1 2 2 2 3	
Locus MC0001 lpxC envA NMC0002 pil51 NMC0003 pil52 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 giycerate dehydrogenase	Sequence length           924           291           366           330           219           954	Genome position 1261 3341 3675 4069 4476 4816	Reference genome 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 2	669 (2846) 1 2 1 2 1 4 2	(2840) 1 2 2 3 2 2	(2844) 1 2 2 5 2 2	(2847) 1 2 2 3 2 2	(FAM18) 1	(2839) 1 2 2 2 3 2 3 2	
Locus MC0001 lpxClenvA NMC0002 pil51 NMC0004 fbp NMC0005 NMC0006 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           954           2058	Genome position 1261 3341 3675 4069 4476 4816 5843	Reference genome 1 1 1	644 (L93/4286) 1 2 2 2 2 2 2 2 2 2 2	662 (2837) 1 2 2 3 2 2 2	665 (2845) 1 2 2 3 2 2 2	666 (2843) 1 2 2 3 2 2 2 2	667 (2842) 1 2 2 3 2 2 2	669 (2846) 1 2 1 2 1 4 2 2 2	(2840) 1 2 2 3 2 2 2 2 2	(2844) 1 2 2 5 2 2 2 2 2	(2847) 1 2 2 3 2 2 2 2 2	(FAM18) 1	(2839) 1 2 2 3 2 3 2 3 3	
Locus NMC0001 lpxClenvA NMC0002 piL51 NMC0003 fb2 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 2	669 (2846) 1 2 1 2 1 4 2	(2840) 1 2 2 3 2 2	(2844) 1 2 2 5 2 2	(2847) 1 2 2 3 2 2	(FAM18) 1	(2839) 1 2 2 2 3 3 2	

# 14.6.3 Include in identifiers fieldset

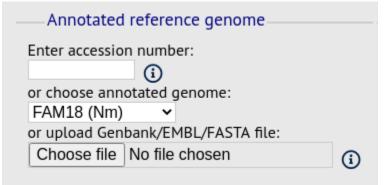
This selection box allows you to choose which isolate provenance fields will be included in the results table. This does not affect the output of the alignments as taxa names are limited in length by the alignment programs.

Include in identifiers	
isolate	<u>^</u>
country	
region	
year	
date sampled	
isoyear sampled	
week sampled	
date received	
non culture	-

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.



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

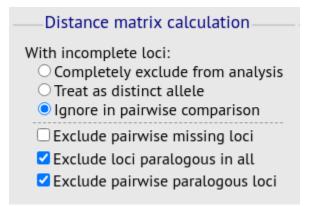
This section allows you to modify BLAST parameters. This affects sensitivity and speed.

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.



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.

	uences by	
Sequence method:	✓ (i)	
Project:		<b>~</b> (i)

Available options are:

- Sequence method Choose to only analyse contigs that have been generated using a particular method. This depends on the method being set when the contigs were uploaded.
- Project Only include isolates belonging to the chosen project. This enables you to select all isolates and filter to a project.

#### 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		
<i>Neisseria</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	
		+
	<b>1</b> INFORMATION	+

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 bigsdbreadthedocsio 12*
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 (uns	pecified/other) Neisseria meningitidis	В	9812 ST-213 complex	22	14	F5-5
19959 M10 240499	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	В	275 ST-269 complex	22	9	F5-12
19960 M10 240500	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	В	41 ST-41/44 complex	7-2	4	F1-5
19961 M10 240502	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	В	340 ST-41/44 complex	7-2	4	F1-5
19962 M10 240503	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	Y	23 ST-23 complex	5-1	2-2	F1-96
19963 M10 240505	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	Y	1655 ST-23 complex	5-1	10-10	F4-1
19964 M10 240507	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	Y	183 ST-23 complex	21	16-5	deleted
19965 M10 240508	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	В	1575	7-2	13-1	F1-7
19966 M10 240511	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	В	4713	22	9	F5-12
19967 M10 240512	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	В	41 ST-41/44 complex	7-1	1	F1-5
19968 M10 240514	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	W	11 ST-11 complex	5	2	F1-1
19969 M10 240515	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	В	269 ST-269 complex	5-1	2-2	F5-1
19970 M10 240520	UK [England] 2010 invasive (uns	pecified/other) Neisseria meningitidis	Y	1655 ST-23 complex	5-1	10-1	F4-1
$\bigcirc \bigcirc 1$							

Analysis tools

5									
Breakdown:	Fields	Two Field	Com	binations	Publications	s Seque	nce bin		
🛃 Analysis:	Codons	Gene Pres	ence	Genome	Comparator	BLAST	rMLST	species id	PCR
Export:	Dataset	Contigs	Sequ	ences					
🗹 Third party:	GrapeTr	ee PhyloV	'iz						

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.

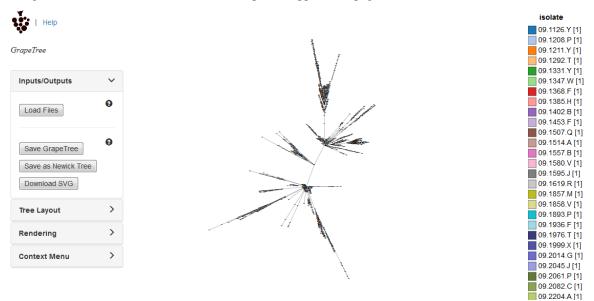
Additional fields can be selected to be included as metadata for use in colouring nodes - select any fields you wish to include. Multiple selections can be made by holding down shift or ctrl while selecting. Click 'Submit' to start the analysis.

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, interesting interesting of the recommended schemes.	
Isolates     Loci     Recommended schemes     Schemes       18969     165_rRNA (SSU_rRNA)     Select one or more schemes     below or use the full schemes list.       19023     165_rRNA (SSU_rRNA)     235_rRNA       19025     abcZ (NEIS1015)     abcZ (NEIS1015)       19026     All None Paste list       Clear     List all	
Include fields     Parameters / options     Action       Select additional fields to include in GrapeTree metadata. <ul> <li>Rescan undesignated loci</li> <li>SUBMIT</li> </ul> Submit           clonal complex (MLST)          cluster (GC OMV peptide typing)          Submit           contrinent          contrivent          contrivent          Submit	

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) MS Tree (Newick format) Metadata (TSV format)
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.

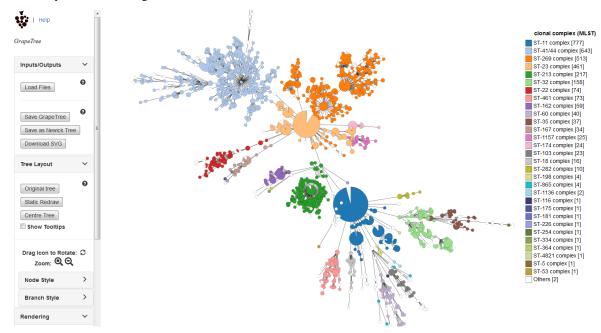


The image can be manipulated in various ways. These include modifying the tree layout, customising node labels and size, modifying branch lengths and collapsing branches. The image can be saved in SVG format which can be further edited in image publishing software such as Inkscape.

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.

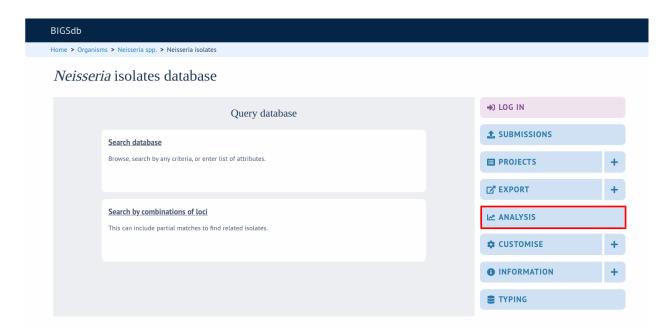
Note: GrapeTree has been described in the following publication:

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.



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 G
Output         Image: Decision of the state 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.

1	/891	B34; NIB5C_2/00; 21034	Finiano	19/2	invasive (unspecified/other)	iveisseria meningiliuis	А	С	SI-5 complex	20	У	1-0-1
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	А	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 > Neisseria	spp. > Neisseria isolates > Plugins > Ir	silico PCR			
In silico PCR					Help 🕜
written by Guy Slater.	reactions run using genomes stored in the			rapper for the exonerate ipcre	ess program
Isolates	ate ids to run the PCR reaction against. Th 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.

lome > Organisms > N	leisseria spp. > Neisseria is	olates > Job status view	er		
ob status v	riewer				
Submit S Start Pro Stop	lob id: BIGSdb_028633_15 time: 2020-07-3112:19:3 itatus: finished time: 2020-07-3112:19:3 gress: 100% time: 2020-07-3112:19: time: 2 seconds	56			
Output           I         A/MI027           1         A/MI027           2         120M           7         7891           10         6748           11         129E           13         139M           19         53131           24         54355           30         14           34         20           35         26           52         2435           64         254           61         393           64         254           62         11-004           84         IAL2229           90         CN100	1 183322 898 1 183655 898 1 183761 898	start         end         descr           6772         7669         forw           1608         2505         rev           6770         7667         forw           1614         2511         rev           1623         2520         rev           1620         2517         rev           6790         7687         forw           1620         2517         rev           6774         7671         forw           983430         984327         forw           1620         2517         rev           6787         7674         forw           1913         2810         rev           6767         7664         forw           1616         2513         rev           6770         7664         forw           1618         2516         rev           6770         7667         forw           7708         8105         forw	vard erse erse erse erse erse erse vard erse erse erse erse erse erse erse er		
	hat summary file	Excel for	mat summary file	FASTA file of predicted product sequences	

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 pro	oducts
Min length:	0
Max length:	10000
🗹 Export sequ	ences

**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 > Organism	ns > Neisseria spp. > Neisseria isolates			
Neisseri	a 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.		MALYSIS	
				+
				+

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 bigsdbreadthedocsio 🗗
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.

.4	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningitidis	Х	117			14	
5	1		Germany	1999	carrier	Neisseria meningitidis	E	864				
16	2		Germany	1999	carrier	Neisseria meningitidis	В	854	ST-18 complex			
17	3		Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (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	invasive (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			
(4	00	1 2 3 4	) <b>&gt;&gt;</b>									
Δ	nalysis	tools										

Reakdown:	Fields	Two Field	Combination	s Polymorphic site	es	Publications	Sequence bin	
🛃 Analysis:	BURST	Codons	Gene Presence	Genome Compar	ator	BLAST	rMLST species id	PCR
🐻 Export:	Dataset	Contigs	Sequences					
🛃 Third party:	GrapeTre	ee iTOL	PhyloViz M	licroreact				

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.

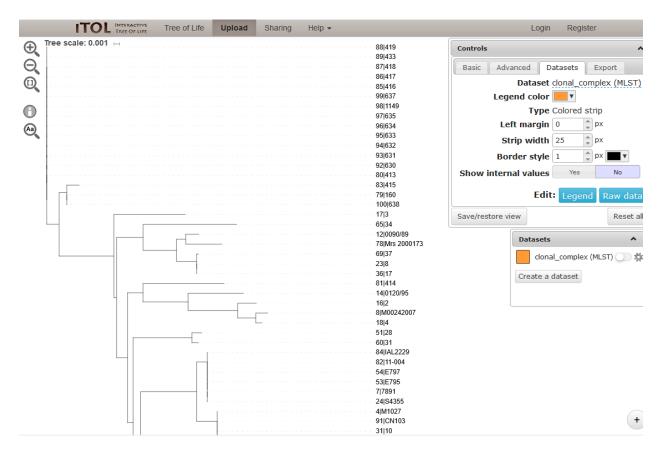
BIGSdb			
Home > Organisms > Neisseria spp	o. > Neisseria isolates > Plugins > iTOL		
iTOL - Interactiv	e Tree of Life - Neisse	ria isolates	Help 🗹
iTOL is develop 1. Biobyte s 2. Europear 3. Max Delt 4. Departm Web site: https://toLembLde/	loads data for analysis within the Interactive Tree bed by: Ivica Letunic (1) and Peer Bork (2,3,4) olutions GmbH, Bothestr 142, 69126 Heidelberg, Molecular Biology Laboratory, Meyerhofstrasse 1 orück Centre for Molecular Medicine, 13125 Berlir ent of Bioinformatics, Biocenter, University of Wür Interactive tree of Life (ITOL) v3: an online tool for	, Germany 1, 69117 Heidelberg, Germany n, Germany rzburg, 97074 Würzburg, Germany	d other trees. <i>Nucleic Acids Res</i> <b>44(W1)</b> ;W242-5.
peptide loci with genome sequence scheme. Analysis is limited to 2,500 record	es, can be included. Please check the loci that you so 110,000 sequences (records x loci).	w would like to include. Alternatively select one o	tabase containing allele sequence identifiers, or DNA and or more schemes to include all loci that are members of the
Isolates	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) 235_rRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) All None Paste list	Recommended schemes below or use the full schemes list. MLST Ribosomal MLST N. meningitidis cgMLST v1.0 N. gonorrhoeae cgMLST v1.0
Schemes Plasmids Typing Finetyping an Server Antigen gener Bessero Antigen Human-restric	azithromycin spectinomycin bioproject accession biosample accession		

The job will be sent to the job queue. When it has finished, the generated tree and associated metadata will be uploaded

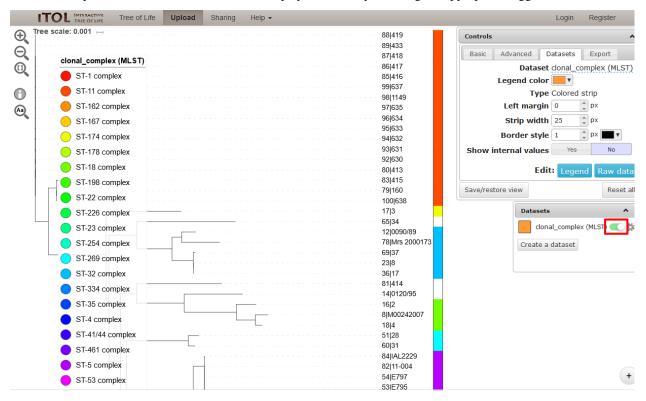
to the Interactive Tree of Life website (https://itol.embl.de/). Click the button marked 'Launch iTOL'.

BIGSdb
Home > Organisms > Neisseria spp. > Neisseria isolates > Job status viewer
Job status viewer
Status         Job id:         BIGSdb_027908_1596206761_04985           Submit time:         2020-07-31 14:46:01           Status:         inished           Stature:         100%           Progress:         100%           Stop time:         2020-07-31 14:46:25           Total 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.

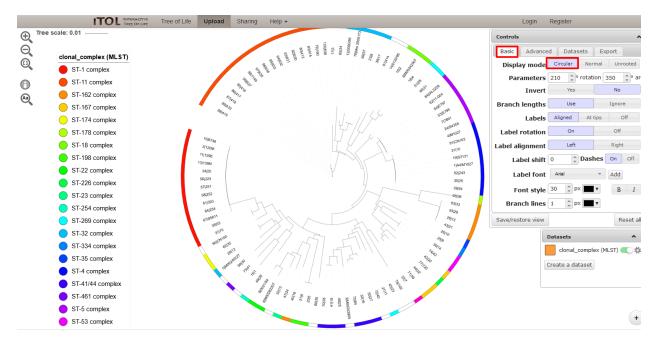
Your browser will open the iTOL website with your tree.



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.

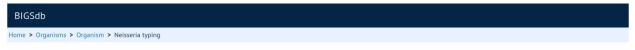


See the detailed documentation on the iTOL website for more information about manipulating and exporting trees.

# 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	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.	Sequence similarity Sequence comparison	
		In a batch	Locus Explorer	
		Look up multiple allelic profiles together.		+
			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 sequences Select analysis	
1       Image: Constraint of the sequence of the seque	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       • Polymorphic Sites - Display polymorphic site frequencies and sequence schematic       SUBMIT         99       • Codon - Calculate G+C content and codon usage       • Translate - Translate DNA to peptide sequences         101       • Translate - Translate DNA to peptide sequences       • SUBMIT         102       • Translate - Translate DNA to peptide sequences       • SUBMIT	
All None	

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.

BIGSdb	
Home > Organisms > Organism > Neisseria typing > Plugins > Locus Explorer	
	Help 🔽
Locus Explorer	
Polymorphic site analysis	
adk	
The colour codes represent the percentage of alleles that have a particular nucleotide at each position. Click anywhere within the sequence to drill down to allele and profile in	formation. The width
of the display can be altered by going to the options page - change this if the display goes off the page.	
100 alleles included in analysis. 112 polymorphic sites found.	
Key: <mark>0 - 1</mark> 0   <mark>&gt; 10 - 20   &gt; 20 - 30   &gt; 30 - 40   &gt;40 - 50   <b>&gt; 50 - 60</b>   <mark>&gt;60 - 70</mark>   <b>&gt;70 - 80</b>   &gt;80 - 90   &gt;90 - 100</mark>	
<u>1 10 20 30 40 50 60 70 80 90 100</u>	
GAAGCGAAAAAAAAATCATTGA <mark>c</mark> gaaggcggittggtgcgcgacgacatcattattgtgcgaagaagaacgcatcgcgcaagaacgactgcgaaaaacggtt 2 T C da ta i 2 T T C ii T G ii T C T T T	
<b>G A</b> 101 110 120 130 140 150 160 170 180 190 200	
TCCTGTTC <mark>GACGGTTCCCCCGCACAT</mark> TGGCACAAGCCGAAGCGATGGTTGAAGCA <mark>GGCGTGGATTTGGAATCGACTC</mark> GTTGAAATCGACTGCCTGACAG	
201 210 220 230 240 250 260 270 280 290 300	
CGTGATTGTCGACCGCATGAGCGGCCGCCGCGGGCATTTGGCTTCCGGCCGTACTTACCACGTTACCACGTCACAAAGTTGAAGGCCAAAAGACGAC T a t t t a t c c a t t c c t t t gatt t gac t t	
A A T A 300 300 300 300 300 300 300 400	
GTAACCGGCGAAGATTTGATTCAGCGCGACGACGACGACAAAGAAGAAGAAACCGTGAAAAAACGCCTTGCCGTTTACCACGAGCAAACCGAAGTTTTGGTCGATT	
AAGC ATTÜ CITGOTIA GOÜG	
401 410 420 430 440 450 460	
TTTACAGCAAACTGGAAGGCGAACACGCCGCCTAAATACATCAAAGTTGACGGCACTCAGCGGGTA C 🗄 A T G T AT 🖯 T 🔍 T T 🔍 GAG 🗛	

Clicking any of the sequence bases will calculate the exact frequencies of the different nucleotides at that position.

- 50   <mark>&gt;50</mark> -	<mark>- 60</mark>   <mark>&gt;60 - 70</mark>   >	<mark>70 - 80</mark>   >80 -	90 >90 - 100	
30	40	50	60	70
GCGG <mark>C</mark> TTG	GT <mark>G</mark> CGCGACGA	CATCA1 T ATC	GGCATGGTCA/	AAGAACG <mark>C</mark> AT
TA 🚺	ATT	C C	Î I	G
		_	G	
130	140	150	160	170

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.



The second part of the page shows a table listing nucleotide frequencies at each of the variable positions.

									_	
		A A	G	С		P		тт	Ū.	
101		41	0		42	0	43	0	44	0
				TGG/					ATACA	
, c			G			ΤG	TA	T G	Т	
			- T.							
ucleo	tide	fre	aue	nci	es					
	aut		140							
					N	lucleoti	de			÷
osition 🕯	·			-						
	A ÷	C ÷	G \$	T ÷					%T \$	%- ÷
6	18	0	82	0		18.00		82.00		
15	0	99	0	1	0		99.00		1.00	
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		
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	
	-			77			(7.00		77.00	

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 — Action	
1     •     Polymorphic Sites - Display polymorphic site frequencies and sequence schematic       2     •     •       3     •     •       3     •     •       4     •     •       5     •       6     •	
All None	

The GC content of the alleles will be determined and a table of the codon frequencies displayed.

IGSdb					
ome > Orga	nisms > O	rganism > Ne	isseria typing	> Plugins > L	xcus Explorer
					Hetp 🗹
locus	Expl	orer			
Codon	Liene	10			
JOUUII	USag	30			
adk					
ORF used: 1					
100 alleles i	included in	analysis.			
GC conter	nt				
Coding: GC 5					
1st letter: GO					
2nd letter: G 3rd letter: G					
Sid tetter: G	C 59.91%				
ord tetter: G	C 59.91%				
ora tetter: G	C 59.91%				
Sid letter: G	C 59.91%				
Codons	C 59.91%				
	C 59.91%				
Codons		usage of a giv	en codon amo	ong its redundar	t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro	oportion of	usage of a giv ven codon per			t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U	oportion of Usage of giv	ven codon per	1000 codons.		t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Codon + A	oportion of Usage of giv Amino acid	<ul> <li>ven codon per</li> <li>Fraction •</li> </ul>	1000 codons. Frequency ¢	Number +	t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Codon • A GCA	oportion of Usage of giv Amino acid A	<ul> <li>Fraction </li> <li>0.265</li> </ul>	1000 codons. Frequency + 17.484	Number ¢	t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Codon • A GCA GCC	oportion of Usage of giv Amino acid A A	<ul> <li>Fraction </li> <li>0.265</li> <li>0.235</li> </ul>	1000 codons. Frequency ◆ 17.484 15.484	Number \$ 271 240	t 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	oportion of Usage of giv Amino acid A A A A	<ul> <li>Fraction </li> <li>0.265</li> <li>0.235</li> <li>0.399</li> </ul>	1000 codons. Frequency ♦ 17.484 15.484 26.323	Number \$ 271 240 408	t 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	oportion of Usage of giv Amino acid A A A A A	<ul> <li>► Fraction ◆</li> <li>● C.265</li> <li>● 0.235</li> <li>● 0.399</li> <li>● 0.101</li> </ul>	1000 codons. Frequency \$ 17.484 15.484 26.323 6.645	Number ◆ 271 240 408 103	t 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 TGC	oportion of Usage of giv Amino acid A A A A A C	<ul> <li>Fraction ◆</li> <li>0.265</li> <li>0.235</li> <li>0.399</li> <li>0.101</li> <li>0.970</li> </ul>	1000 codons. Frequency ♦ 17.484 15.484 26.323 6.645 6.323	Number ♦ 271 240 408 103 98	t 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 TGT	oportion of Usage of giv Mino acid A A A A C C C	<ul> <li>► Fraction ◆</li> <li>0.265</li> <li>0.235</li> <li>0.399</li> <li>0.101</li> <li>0.970</li> <li>0.030</li> </ul>	1000 codons. Frequency \$ 17.484 15.484 26.323 6.645 6.323 0.194	Number ◆ 271 240 408 103 98 3	t 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 GCG GCG GCG GCC TGT GAC	poportion of Usage of giv Mino acid A A A A A C C C D	<ul> <li>► raction ◆</li> <li>► Fraction ◆</li> <li>0.265</li> <li>0.235</li> <li>0.399</li> <li>0.101</li> <li>0.970</li> <li>0.030</li> <li>0.745</li> </ul>	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	t 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 GCT TGC TGC TGT GAC GAT	oportion of Usage of giv Mino acid A A A A A C C C D D	<ul> <li>▶ Fraction ◆</li> <li>▶ Fraction ◆</li> <li>0.265</li> <li>0.235</li> <li>0.399</li> <li>0.101</li> <li>0.970</li> <li>0.030</li> <li>0.745</li> <li>0.255</li> </ul>	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	t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: U Codon e A GCA GCC GCC GCC GCC GCC GCC GCC GCC GC	oportion of giv Usage of giv A A A A A C C C C D D D E	<ul> <li>ven codon per</li> <li>Fraction ◆</li> <li>0.265</li> <li>0.235</li> <li>0.399</li> <li>0.101</li> <li>0.970</li> <li>0.030</li> <li>0.745</li> <li>0.255</li> <li>0.924</li> </ul>	1000 codons. Frequency	Number   271 240 408 103 98 3 1412 484 1292	t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Codon • A GCA GCC GCT GCC GCT GCC GAC GAC GAC GAC GAA GAG	poportion of Usage of giv A A A A A C C C C D D D E E	<ul> <li>✓ Fraction ◆</li> <li>✓ Fraction ◆</li> <li>✓ 0.265</li> <li>✓ 0.235</li> <li>✓ 0.399</li> <li>✓ 0.101</li> <li>✓ 0.970</li> <li>✓ 0.030</li> <li>✓ 0.745</li> <li>✓ 0.255</li> <li>✓ 0.924</li> <li>✓ 0.076</li> </ul>	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	t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Prc Frequency: L Codon • A GCA GCC GCC GCC GCC GCC TGC GAT GAA GAA GAA GAA GAA	aportion of Usage of giv Amino acid A A A A C C C D D E E E E F	<ul> <li>► Fraction ●</li> <li>0.265</li> <li>0.235</li> <li>0.399</li> <li>0.101</li> <li>0.970</li> <li>0.030</li> <li>0.745</li> <li>0.255</li> <li>0.924</li> <li>0.076</li> <li>0.614</li> </ul>	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	t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Pro Frequency: L Godon 9 A GCA GCC GCC GCC GCC GCC GCC GCC GCC GC	portion of Usage of giv Amino acid A A A A A C C D D D D D D E E E F F	<ul> <li>► Fraction ►</li> <li>0.265</li> <li>0.235</li> <li>0.399</li> <li>0.101</li> <li>0.970</li> <li>0.030</li> <li>0.745</li> <li>0.255</li> <li>0.924</li> <li>0.076</li> <li>0.614</li> <li>0.386</li> </ul>	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	t set (i.e. the set of codons which code for this codon's amino acid).
Codons Fraction: Prc Frequency: L Codon • A GCA GCC GCC GCC GCC GCC TGC GAT GAA GAA GAA GAA GAA	aportion of Usage of giv Amino acid A A A A C C C D D E E E E F	<ul> <li>► Fraction ●</li> <li>0.265</li> <li>0.235</li> <li>0.399</li> <li>0.101</li> <li>0.970</li> <li>0.030</li> <li>0.745</li> <li>0.255</li> <li>0.924</li> <li>0.076</li> <li>0.614</li> </ul>	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	t 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       Opolymorphic Sites - Display polymorphics its frequencies and sequence schematic       SUBM         2       Codon - Calculate G+C content and codon usage       SUBM         3       Translate - Translate DNA to peptide sequences       SUBM         Att       None	

An aligned amino acid sequence will be displayed.

Home > Or	isms > Organism > Neisseria typing > Plugins > Locus Explorer	
Locus	Explorer	Help 🗹
Trans	te - aligned protein sequences	
adk		
auk		
ORF used:		
The width	the alignment can be varied by going to the options page.	
100 allele	cluded in analysis.	
200 0000		
	10 20 30 40 50 60 70 80 90 100	
1_1	:	
$1_{2_{1}}^{1_{1}}$	:	
1_1 2_1 3_1	:	
1_1 2_1 3_1 4_1	:	
1_1 2_1 3_1 4_1 5_1 6_1	:  :  :  :  :  :  :  :  :	
1_1 2_1 3_1 4_1 5_1 6_1 7_1		
1_1 2_1 3_1 4_1 5_1 6_1 7_1 8_1		
$1_{2} 1 \\ 2_{1} \\ 3_{1} \\ 4_{1} \\ 5_{1} \\ 6_{1} \\ 7_{1} \\ 8_{1} \\ 9_{1} $		
10 1		
$1\overline{0}_1$ 11 1		
$10^{-1}$ 11_1 12_1		
$10^{-1}$ 11_1 12_1 13_1	К	
$10^{-1}$ $11^{-1}$ $12^{-1}$ $13^{-1}$ $14^{-1}$	к	
$10^{-1}$ $11^{-1}$ $12^{-1}$ $13^{-1}$ $14^{-1}$ $15^{-1}$	К	
$10^{-1}$ $11^{-1}$ $12^{-1}$ $13^{-1}$ $14^{-1}$	к	

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	DILOG IN	
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		
This can include partial matches to find related isolates.	CUSTOMISE	+
This can include partial matches to find related isolates.	CUSTOMISE	+
This can include partial matches to find related isolates.		

Jump to the 'Third party' category, follow the link to BLAST, then click 'Launch Microreact'.

Microreact
Summary: Open data visualization and sharing for genomic epidemiology
Microreact is a tool for visualising genomic epidemiology and phylogeography (Argimón et al 2016 Microb Genom 2:e000093). 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.
Documentation bigsb.readthedocs.io 🗗
Launch 'Microreact'

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	А	4	ST-4 complex	7	13	F1-5	
61	393	B392; NIBSC 2823; Z1392	Greece	1968	carrier	Neisseria meningitidis	Α	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	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	Α	21		5-2	10	F3-9	
120	F4698	NIBSC_2731; Z3515	Saudi Arabia	1987	carrier	Neisseria meningitidis	Α	5	ST-5 complex	20	9	F3-1	
128	F6124	NIBSC_2730; Z3524	Chad	1988	invasive (unspecified/other)	Neisseria meningitidis	Α	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	Α	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 🕡 🔊											
Ar	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.

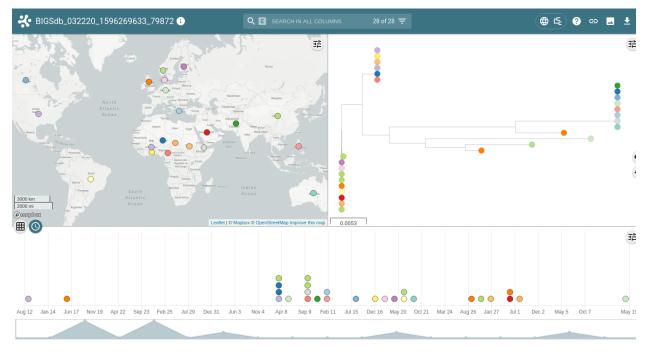
Additional fields can be selected to be included as metadata for use in colouring nodes - select any fields you wish to include. Multiple selections can be made by holding down shift or ctrl while selecting. Click 'Submit' to start the analysis.

BIGSdb		
Home > Organisms > Neisseria spp. > Neisseria isolates > Plugins > Microreact		
Microreact - Neisseria isolates	Help 🖌	1
This plugin uploads data for analysis within the Microreact online service: Microreact is developed at the The Centre for Genomic Pathogen Surveillance by a team Web site: https://microreact.org Publication: Argimón at al. (2016) Microreact: visualizing and sharing data for genomic e		
	Only DNA loci that have a corresponding database containing allele sequence identifiers, or DNA and uld like to include. Alternatively select one or more schemes to include all loci that are members of the	
	Recommended schemes       Schemes         Select one or more schemes       Below or use the full schemes list.         MLST       MLST         Ribosomal MLST       MLST         N. meningitidis cgMLST v1.0       MLST         Clear       Interpring antigen genes         Image: Clear       MLST	
Descriptions	Include fields Action	
Modify the values below - these will be displayed within the created Microreact project Title: Description: Neisseria isolates	ct. Select additional fields to include in Microreact data table- (isolate, country and year are always included).	

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           Job id:         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.

BIGSdb			
Home > Organisr	is > Neisseria spp. > Neisseria isolates		
Neisser	a 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.	MALYSIS	
		CUSTOMISE	+
		<b>1</b> INFORMATION	+
		S TYPING	

Jump to the 'Third party' category, follow the link to PhyloViz, then click 'Launch PhyloViz'.

# <text><text><text><text>

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.

.5 1													
			Germany	1999 0	carrier	Neisseria meningit	idis	E	864				
.6 2			Germany	1999 0	arrier	Neisseria meningit	idis	В	854	ST-18 complex			
7 3			Germany	1999 0	carrier	Neisseria meningit	idis	W	174	ST-174 complex			
.8 4			Germany	1999 0	carrier	Neisseria meningit	idis	В	19	ST-18 complex			
9 S3131	B213; NIBSC_2	813; Z1213	Ghana	1973 invasive (ur	nspecified/other)	Neisseria meningit	idis	A	4	ST-4 complex	7	13-1	F1-
0 5			Germany	1999 0	carrier	Neisseria meningit	idis	NG	198	ST-198 complex			
1 6			Germany	1999 0	carrier	Neisseria meningit	idis	NG	198	ST-198 complex			
2 7			Germany	1999 0	arrier	Neisseria meningit	idis	E	60	ST-60 complex			
3 8			Germany	1999 0	carrier	Neisseria meningit	idis	В	32	ST-32 complex			
4 S4355	B227; NIBSC_2	806; Z1227	Denmark	1974 invasive (ur	nspecified/other)	Neisseria meningit	idis	A	5	ST-5 complex	5-1	9	F3-
5 9			Germany	1999 0	arrier	Neisseria meningit	idis	В	930	ST-334 complex			
(*)     1     2     3     4     5     6     (>)     (>)   Analysis tools													
(() () Analysis	1 2 3 tools	456	00	)									
(() Analysis () Break		4 5 6 Two Field	Combinations	Polymorphic site	es Publications	Sequence bin							
5	down: Fields	4 5 6 Two Field Codons	Combinations Gene Presence	Polymorphic site Genome Compar		Sequence bin rMLST species id	PCR						

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.

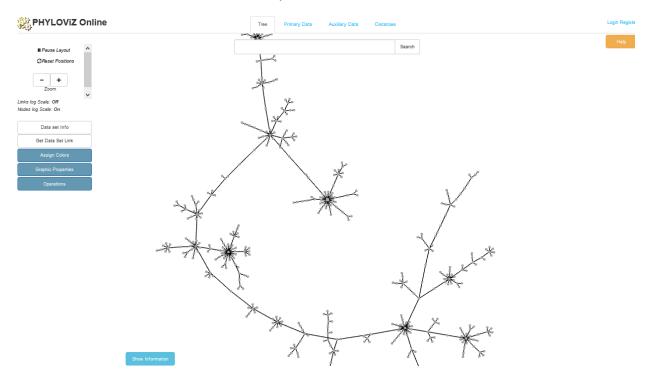
Provenance fields can be selected to be included as metadata for use in colouring nodes - select any fields you wish to include. Multiple selections can be made by holding down Shift or Ctrl while selecting. Click 'Submit' to start the analysis.

BIGSdb				
Home > Organisms > Neisseria spp. > Neisseria is	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 Técnico	a, Instituto de Medicina Molecula 9, 1000-029 Lisboa, Portugal 5, Universidade de Lisboa, Lisboa,	iré Carriço (1), Alexandre P. Francisco (2,3), Càtia Vaz r, Faculdade de Medicina, Universidade de Lisboa, L		
		our nodes based on field values. Analysis is limited t		
Select ids Paste in list of ids to include, start a new line for each. Leave blank to include all ids.	Provenance fields id isolate country continent region year date sampled isoyear sampled	Loci "ISS_rDNA ISS_rRNA (SSU_rRNA) 23S_rRNA abc2 abc2 (NEIS1015) aceF (NEIS1279) All None Paste list	Schemes	×

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

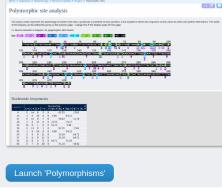
The Polymorphisms plugin generates a *Locus Explorer* polymorphic site analysis on the alleles designated in an isolate dataset following a query.

The analysis can be accessed by selecting the 'Analysis' section on the main contents page.

BIGSdb				
Home > Organisr	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	+
			C EXPORT	+
	Search by combinations of loci This can include partial matches to find related isolates.		MALYSIS	
			CUSTOMISE	+
			<b>1</b> INFORMATION	+

Jump to the 'Breakdown' category, follow the link to Polymorphisms, then click 'Launch Polymorphisms'.

# Polymorphisms Summary: Tool for analysing polymorphic sites for particular locus in an isolate dataset This plugin generates a schematic of the selected locus showing all the polymorphic sites present in the selected dataset. These are also shown in a tabular form with precise frequencies for each nucleotide at every position. Documentation Digidibreadthedocsio C



The analysis is accessed by clicking the 'Polymorphic sites' button in the Breakdown list at the bottom of a results table following a query.

	1		Germany	1333	carrier	Neissena meningitiois	E	804				
5	2		Germany	1999	carrier	Neisseria meningitidis	В	854	ST-18 complex			
7	3		Germany	1999	carrier	Neisseria meningitidis	W	174	ST-174 complex			
8	4		Germany	1999	carrier	Neisseria meningitidis	В	19	ST-18 complex			
.9	S3131	B213; NIBSC_2813; Z1213			,	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					
2	7		Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex			
3	8		Germany	1999	carrier	Neisseria meningitidis	В	32	ST-32 complex			
4	S4355	B227; NIBSC_2806; Z1227	Denmark		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	1 2 3 4	<b>&gt; &gt;</b>									
(« Ar	alysis 1		00									
	alysis 1		Combinations	Polymo	orphic sites Publication	s Sequence bin						
	alysis 1	tools own: Fields Two Field	Combinations Gene Presence		orphic sites Publication e Comparator BLAST	s Sequence bin rMLST species id PCR	E I					
	alysis f	tools own: Fields Two Field ysis: BURST Codons					L					

Select the locus that you would like to analyse from the list.

BIGSdb			
Home > Organisms > Neisseria spp. > Neisseria i	solates > Plugins > Polymorphic sites		
Polymorphic site analys	is	G	Help 🕜 🕒
		few minutes (or longer depending on the queue). This is because sequ lon. Action	iences
'16S_rDNA 16S rRNA (SSU rRNA) 23S_rRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) ackA2 (NEIS1727) acnA (NEIS1729)	If both allele designations and tagged sequences exist for a locus, choose how you want these handled: Use allele sequence retrieved from external database Use sequences tagged from the bin Analyse single example of each unique sequence Exclude incomplete sequences	ANALYSE	

#### 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.

																		Help 📝
Polyn	nor	ph	ic	sit	e a	ana	lysi	S										
											otide at each po blay goes off th	osition. Click anyw e page.	here within th	ne sequence	to drill dow	to allele and	profile informat	tion. The widt
17 alleles	includ	ed in	analy	/sis. 8	1 pc	lymorp	hic site	es found	d.									
Key: <mark>0 - 1</mark> 0	0   <mark>&gt;10</mark> ·	- 20	>20 -	30	>30 -	40   >4	0 - 50	>50 - 6	<mark>60</mark>   >60 - 70	>70 - 80	>80 - 90 >90 -	100						
	1		LO		2			0	40	50	60	70	80	90	100			
				GCCG								TCAGCCATGAGTT						
4.0					4.0	A				A	CCCC		GC TAA	AAAC (	TC A 200			
10		1 GAA		ACGA	12 ATT		13 I GAA		140 AGCGAAGG			170 GGCAGT <mark>CAAGCAG</mark>		190 SAACTCGGTT				
	A	G	T	T		G	CCG			T	AT	G G A	G	T	A			
20	14	~	0		22	0	23	0	240	250	260	270	180	290	300			
26			LO NAATI	CGGC					240 AAAAAGCG	250 TGTCGCCTT		TGGGTGCAGAAGC	280 CCGACGTAT					
	ū					С		A		C T TC	AT	A	ТТ	C	T G			
30			L0	_	32		33		340	350	360	370	380	390	400			
	ACCAA	CCA	IIG				ATTAI	IGG	IGGAAAA	CIGCICAA	GCGIIIGAAG	GCAGCTTGGT	GATTACCCAU	GALLGLLG	IIIIGG			
						<b>T</b>								T	0			
40	91	4	10	С	42		43	0		Ū G		C		I	C			
46			LO BCCA		42		43 AACTC							I	C			
46			GCCA	CGCG	42 GAT	0	AACT					2 2		I	C			
46	A <mark>C</mark> AAT	ATC	GCCA	CGCG	42 GAT	о ТСТСС	AACT	GATC		T		•		I	C			
46	A <mark>C</mark> AAT	ATC	GCCA	CGCG	42 GAT	о ТСТСС	AACT	GATC				0		T	C			
	ACAAT T	ATCO T	GCCA	ecce A	42 GAT	о ТСТСС	AACT	GATC	: 0			00		T	C			
40 Nucleo	ACAAT T	ATCO T	GCCA	ecce A	42 GAT	O TGTCG	AACT	GATC	1 1			2 2		T	C			
	acaat I otide	fre	que	A A	42 GAT C	0 TGTCG C T	de	GATC		1		3 3		T	c			
Nucleo	ACAAT	fre	que	ncie	42 GAT C es N	0 TGTCG C T	de %C ÷	GATC		1		3 3		I	C			
Nucleo Position 9	ACAAT	fre	que G ÷	ncie	42 GAT C C S N - + 0	0 TGTCG C T ucleotic	de	GATC C	%T <b>≎</b> % 17.65	1		3 3		I	C			
Nucleo	ACAAT	fre	que	ncie	42 GAT C es N	0 TGTCG C T	AACTC T de %C \$ 82.35	GATC	17.65	1		3 3		I	C			
Nucleo Position 9 24	ACAAT otide A + 0 1	fre	que G ÷ 0 16	ncie T ÷	42 GAT C C N - + 0 0	0 TGTCG C T ucleotic	de %C ÷	GATC C		1		3 3		I	C			
Nucleo Position 9 24 27 28 30	ACAAT ■ otide	ATCC T fre C ÷ 14 0 10 0 0	G ¢ 0 16 0 13 1	<b>CECCE</b> A ncie T + 3 0 7 0 0	422 GAT C C S N - ¢ 0 0 0 0 0 0 0 0	0 TGTCG C T ucleotic %A ÷ 5.88	AACTC de %C ¢ 82.35 58.82	<b>GATC</b> <b>C</b> 94.12	17.65 41.18	1		3 3		I	c			
Position 9 24 27 28 36	ACAAT ■ otide	ATCC T fre C ↓ 14 0 0 0 14	G € 0 16 0 13 1 0	CCCCC A ncie 3 0 7 0 0 3	422 GAT C C C C C C C C C C C C C C C C C C C	0 TGTCG C T UCLeotic %A ¢ 5.88 23.53 94.12	AACTC T de %C \$ 82.35	GATC € 94.12 76.47 5.88	17.65	1		3 3		T	ç			
Position 9 24 27 28 30 36 45	ACAAT ↓ A ↓ 0 1 0 4 16 0 1	ATCC T fre 14 0 10 0 14 0	G ¢ 0 16 0 16 16	CCECE A ncie 3 0 7 0 0 3 0 0 3 0	42 GAT C C N - ↓ 0 0 0 0 0 0 0 0 0 0 0 0 0	0 TGTCG C T UCLEOTIO %A \$ 5.88 23.53	AACTO de %C \$ 82.35 58.82 82.35	GATC С 94.12 76.47	17.65 41.18 17.65	1		3 3		T	ç			
Position 9 24 27 28 30 36 45 54	ACAAT Coticle Cotic	ATCC I fre C ↓ 14 0 0 0 14 0 6	G ≠ 0 16 0 16 0 16 0	CCECCE A ncie 3 0 7 0 0 3 0 11	422 GAT C C C C C C C C C C C C C C C C C C C	0 TGTCG C T UCLeotic %A ¢ 5.88 23.53 94.12	de %C \$ 82.35 58.82 82.35 35.29	GATC € 94.12 76.47 5.88	17.65 41.18 17.65 64.71	1		3 3		T	ç			
Position 9 24 27 28 30 36 45	ACAAT ↓ A ↓ 0 1 0 4 16 0 1	ATCC T fre 14 0 10 0 14 0	G ¢ 0 16 0 16 16	CCECE A ncie 3 0 7 0 0 3 0 0 3 0	42 GAT C C N - ↓ 0 0 0 0 0 0 0 0 0 0 0 0 0	0 TGTCG C T UCLeotic %A ¢ 5.88 23.53 94.12	AACTO de %C \$ 82.35 58.82 82.35	GATC € 94.12 76.47 5.88	17.65 41.18 17.65	1				T	C			
Position 9 24 27 28 30 36 45 54 57	ACAAT Cotide	ATCC T fre C ↓ 14 0 10 0 14 0 6 6	G ≠ 0 16 0 13 1 0 16 0 0	<b>CCCCC</b> A ncie 3 0 7 0 0 3 0 11	42 GAT C C C C C C C C C C C C C C C C C C C	0 TGTCG C T UCLeotic %A \$ 5.88 23.53 94.12 5.88	de %C \$ 82.35 58.82 82.35 35.29	GATC C 94.12 76.47 5.88 94.12	17.65 41.18 17.65 64.71	1				T	<u>c</u>			
Nucleo Position 9 24 27 28 30 36 45 54 57 59 60 60 63	ACAAT 0 0 0 1 0 1 0 1 0 1 0 1 0 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 0 0 0 0 0 0 0 0 0 0	Arcc T free C ↓ 14 0 0 14 0 0 14 0 6 6 0 7 16	G ≠ 0 16 0 13 1 0 16 0 0 6 0 0	CCCCC A ncie 3 0 7 0 0 3 0 11 11 0 10 1	42 GAT C C N C C C C C C C C C C C C C C C C	0 TGTCG T Vucleotid %A \$ 5.88 23.53 94.12 5.88 64.71	de %C ↓ 82.35 58.82 82.35 35.29 35.29	GA1C C 94.12 76.47 5.88 94.12 35.29	17.65 41.18 17.65 64.71 64.71	1				T	c			
Nucleo Position 9 24 27 28 30 36 45 57 59 60 63 364	ACAAT	Arcc T fre C + 14 0 10 0 0 14 0 6 6 0 7 16 0	G ¢ 0 16 0 13 1 0 16 0 0 6 0 0 6	CGCC A ncie 3 0 7 0 0 3 0 0 11 11 0 10 1 0	42 GAT C C N C C C C C C C C C C C C C C C C	0 IGTCG C I UCLeotic %A ¢ 5.88 23.53 94.12 5.88 64.71	de %C + 82.35 58.82 82.35 35.29 35.29 41.18 94.12	GATC C 94.12 76.47 5.88 94.12	17.65 41.18 17.65 64.71 64.71 58.82 5.88	1				T	C			
Nucleo Position 9 24 27 28 30 36 45 54 554 554 554 60 63 64 66	ACAAT T t t t t t t t t t t t t t	Arcc T fre C ← 14 0 10 0 0 14 0 6 6 0 7 16 0 13	G ¢ 0 16 0 13 1 0 16 0 0 6 0 0 6 0 0	CGCCC A ncie 3 0 7 0 0 3 0 11 11 0 10 10 10 10 4	42 GAT C C N C C C C C C C C C C C C C C C C	0 TGTCG T Vucleotid %A \$ 5.88 23.53 94.12 5.88 64.71	AACTC de %C + 82.35 58.82 82.35 35.29 41.18	GA1C C 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	1				T	C			
Nucles Position 9 24 27 28 30 36 45 57 59 60 63 64 66 72	ACAAT A + 0 1 0 4 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0	Arcc T fre 14 0 10 0 14 0 6 6 0 7 16 0 13 0	G ♦ 0 16 0 13 1 0 0 6 0 0 6 0 0 14	CCCCC A ncia 3 0 7 7 0 0 3 0 11 11 0 10 10 10 1 0 4 3	42 GAT C C N C C C C C C C C C C C C C C C C	0 TGTCG C T %A ↓ 5.88 23.53 94.12 5.88 64.71 64.71	de %C + 82.35 58.82 82.35 35.29 35.29 41.18 94.12	<ul> <li>GATC</li> <li>C</li> <li>%G ◆</li> <li>94.12</li> <li>76.47</li> <li>5.88</li> <li>94.12</li> <li>35.29</li> <li>35.29</li> <li>82.35</li> </ul>	17.65 41.18 17.65 64.71 64.71 58.82 5.88	1				T	C .			
Nucleo Position 9 24 27 28 30 36 45 54 554 554 554 60 63 64 66	ACAAT T t t t t t t t t t t t t t	Arcc T fre C ← 14 0 10 0 0 14 0 6 6 0 7 16 0 13	G ¢ 0 16 0 13 1 0 16 0 0 6 0 0 6 0 0	CGCCC A ncie 3 0 7 0 0 3 0 11 11 0 10 10 10 10 4	42 GAT C C N C C C C C C C C C C C C C C C C	0 TGTCG T Vucleotid %A \$ 5.88 23.53 94.12 5.88 64.71	de %C \$ 82.35 58.82 82.35 55.29 41.18 94.12 76.47	GA1C C 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	1				T	c			
Position 9 24 27 28 30 36 45 54 57 59 60 63 64 66 63 64 66 72 78	ACAAT	Arcc T free 14 0 10 0 14 0 6 6 0 7 16 0 13 0 0	G ♦ 0 16 0 13 1 0 0 6 0 0 6 0 0 6 0 0 14 2	CCCCC A ncia 3 0 7 7 0 0 3 0 11 11 0 10 1 0 4 3 0	42 GAT C C N - ↓ O O O O O O O O O O O O O O O O O O O	0 TGTCCC T UCLEOTIO %A + 5.88 23.53 94.12 5.88 64.71 64.71 88.24	de %C \$ 82.35 58.82 82.35 55.29 41.18 94.12 76.47	<ul> <li>GATC</li> <li>C</li> <li>%G ◆</li> <li>94.12</li> <li>76.47</li> <li>5.88</li> <li>94.12</li> <li>35.29</li> <li>35.29</li> <li>82.35</li> </ul>	17.65 41.18 17.65 64.71 64.71 58.82 5.88 23.53	1				T	c			

# 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 > Organisms > Neisseria spp. > Neisseria isolates	
<i>Neisseria</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.	Le ANALYSIS
This can include particle narelies to this reaced isolates.	¢ CUSTOMISE +
	INFORMATION     +

Jump to the 'Analysis' category, follow the link to rMLST species identity, then click 'Launch rMLST species identity'.

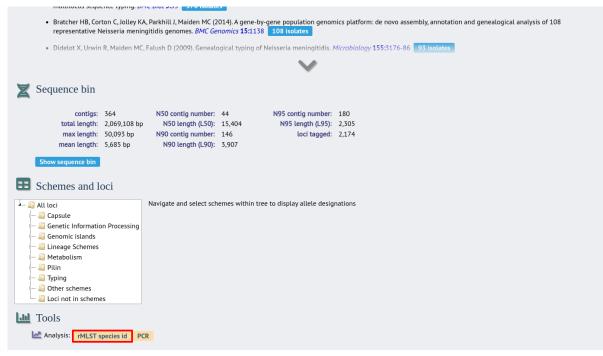
The tool can be accessed from the front page of an isolate database.

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 bigsdbreadthedocsio 😋
Normality       Normality
Launch 'rMLST species identity'

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		len	ce bin size >= 2	Mbp 🗸	€ € Add	
		r by: id olay: 25 ✓ reco	ords per pa	ge 🚺		✓ ascending	* RESET	SEARCH					
7	records retu	irned. Click the h	yperlinks fo	or detailed i		tion. te fields (j)				MLST	Fine	typing anti	
id	isolate	aliases		country	vear		species	capsule group	ST	clonal complex			
	A4/M1027	B1; NIBSC 280	-	USA		invasive (unspecified/other)					5-2	10	F1-5
2	120M	B35; NIBSC_282	22; Z1035	Pakistan		meningitis and septicaemia			1	ST-1 complex	5-2	10	F5-1
7	7891	B54; NIBSC_276		Finland		invasive (unspecified/other)				ST-5 complex	20	9	F3-1
0	6748	B73; NIBSC_278				invasive (unspecified/other)				ST-1 complex	18-1	3	F5-1
1		B92; NIBSC_282				invasive (unspecified/other)				ST-1 complex	5-2	10	F3-6
3		B99; NIBSC_279 B213; NIBSC_28				invasive (unspecified/other)	Neisseria meningitidis			ST-1 complex ST-4 complex	5-2 7	10 13-1	F5-1 F1-5
	nalysis		10,1110	Chana	1775		Tresseria meningreais	A		51 Complex		101	
	🕓 Breakd	down: Fields	Two Field	Combina	tions	Polymorphic sites Public	cations Sequence bin						
	🛃 Ana	-	Codons	Gene Pres	ence	Genome Comparator BLA	AST rMLST species id	PCR					
	🛃 🗠 Ana	-		Gene Pres		Genome Comparator BLA	AST 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 s	op. > Neisseria isolates > Plugins > Spe	cies identification		
rMLST species	identification			Help 🕜 🟮
validated genomes in the rMLS			l be shown where identified alleles have	been linked to
	Options	Action		
1 2 7 10 11 13 19 Clear List all	<ul> <li>Scan genome or use allele designations already stored</li> <li>Scan genomes</li> <li>Use stored allele designations</li> </ul>	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.

I.				Prediction from identified rMLST alleles linked to genomes			Identified rSTs
	isolate	Rank	Taxon	Taxonomy	Support Match	es 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 (IS	ON format)				

Clicking the '+' icon on any row will display further details about the matches.

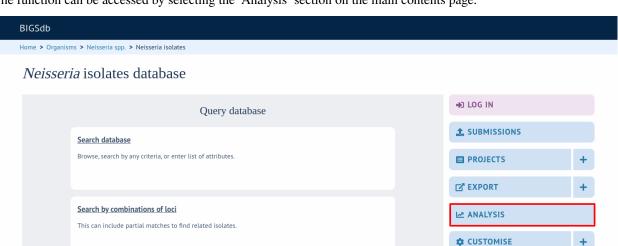
	Inclote						Prediction from identified rMLST alleles linked to genomes			Identified rSTs
id	isolate	Rank	(	Тахо	n		Taxonomy Support Mate	hes	rST	Species
1	A4/M1027	SPECI		Neisse meningi		Proteoba meningit	cteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria 100% C		2504	Neisseria meningitidis
2	120M	SPECI		Neisse meningi		Proteoba meningit	teria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria dis		2532	Neisseria meningitidis
7	7891	SPECI		Neisse meningi		Proteoba meningit	rteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria 100% C		2434	Neisseria meningitidis
10	6748	SPECI		Neisse meningi		Proteoba meningit	rteria > Betaproteobacteria > Neisseriales > Neisseriaceae > Neisseria > Neisseria  100%		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			
BA	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	519	181474		11001	species: Neisseria meningitidis [n=17221]; Neisseria gonorrhoeae [n=1]; Neisseria sp. [n=1]			
BA	CT000006	5	369	181603	53216	53584	species: Neisseria meningitidis [n=2053]; Neisseria sp. [n=1]			
BA	CT000007	1	471	181145	2253	2723	species: Neisseria meningitidis [n=14753]; Neisseria sp. [n=1]			
BA	CT000007	1	471	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	20941	21333	species: Neisseria meningitidis [n=3125]			
	CT000010	1		181315		20144	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'.

**1** INFORMATION

The function can be selected by clicking the 'Sequence bin' link on the Breakdown section of the main contents page.

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 latter letric 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.
ocumentation bigsdb.readthedocs.io 🗹
icida na sigan sikana a kana kana kana kana kana kana rakadawa nd sequence bin contig properties
And the specific transmitter works of the specific transmitter specific
Important
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.

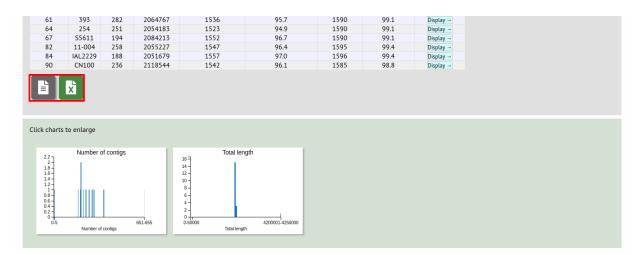
BIGSdb			
Home > Organisms > Neisseria sp	p. > Neisseria isolates > Plugins > Sequence bi	n breakdown	
Breakdown of se	equence bin contig prop	perties	Help 🗹 🗿
Please select the required isolate	ids for comparison - use Ctrl or Shift to make mul	tiple selections. Select loci/schemes to use	for calculating percentage of alleles designated or tagged.
Isolates	Loci	Recommended schemes	Schemes
1 2 7 10 11 13 10 Clear List all	165_rDNA 165_rRNA (SSU_rRNA) 235_rRNA abcZ abcZ (NEIS1015) accF (NEIS1279) All None Paste list	Select one or more schemes below or use the full schemes list. MLST R. meningitidis cgMLST v1.0 N. gonorrhoeae cgMLST v1.0 CLear	Wexsero Antigen Sequence     With the serie     With the serie     With the serie     With the serie serie     With the series of the ser
Options	Filter by		Action
☐ Contig analysis (min, max, M ☐ Calculate %GC	N50 etc.) Sequence method: Project: Experiment:	• 0 • 0	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.

	lear List all		'16S_rDNA 16S_rRNA (1 23S_rRNA abcZ abcZ (NEIS1 aceF (NEIS1	015) 279) I None Paste list	Ţ	MLST Ribosomal ML N. meningitidi N. gonorrhoea	ne full schemes list.		Genom Genom Genom Genom Genom Genom Genom Pilin Plasmic	: Information Process ic islands e Schemes ilism prrhoeae AMR	sin•
Option	ns		Fi	lter by				— Actio	on		
	ig analysis (min, ı ılate %GC	max, N50	etc.)	Sequence method: Project: Experiment:	<b>`</b> (j)	<b>`</b> (j)	<b>~</b> (j	RES	SET SUBM	ШТ	
Loci selecte	ed: 1605										
Isolate id a	+ Isolate + Cor	ntias + 1	otal length +	Alleles designated +	% Alleles desi	onated + Loci t	agged ÷ % Loci tag	aaed ‡	Sequence bin +		
1		364	2069108	1516	94.5	-	591 99.1		Display -		
2	120M	359	2059411	1523	94.9	1	593 99.3	3	Display -		
7	7891	199	2057385	1565	97.5	1	594 99.3	3	Display →		
10	6748	652	4241338	1566	97.6	1	590 99.1	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:	<b>~</b> (j		0	
	- + 1 ( 05							
select	ed: 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

Query database	DI 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				
This can include partial matches to find related isolates.		+		
This can include partial matches to find related isolates.		+		
This can include partial matches to find related isolates.		++		

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.readthedocs.io 🗗
BCCCM Here a sequence is housen and is hous
Breakdown of capsule_group by closal_complex (MLST): Interaction Data was eta An The Anti- Reg State parameters Interaction Data was and the Anti- Interaction Data was a
Cains Calanstop
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.

12	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningiti	dis	В	1015	ST-32 complex	7	16	
.3	139M	B99; NIBSC_2795; Z1099	Philippines	1968		Neisseria meningitie	dis	A	1	ST-1 complex	5-2	10	F5-1
	0120/95		Czech Republic	1995	invasive (unspecified/other)	Neisseria meningiti	dis	Х	117			14	
.5	1		Germany	1999	carrier	Neisseria meningiti	dis	E	864				
16	2		Germany	1999	carrier	Neisseria meningiti		В	854	ST-18 complex			
.7	3		Germany	1999	carrier	Neisseria meningiti		W	174	ST-174 complex			
.8	4		Germany	1999	carrier	Neisseria meningiti		В	19	ST-18 complex			
.9	S3131	B213; NIBSC_2813; Z1213	Ghana		invasive (unspecified/other)			A	4	ST-4 complex	7	13-1	F1-5
0	5		Germany	1999	carrier	Neisseria meningiti		NG		ST-198 complex			
1	6		Germany	1999	carrier	Neisseria meningiti		NG		ST-198 complex			
2	7		Germany	1999	carrier	Neisseria meningiti		E	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningiti		В	32	ST-32 complex			
	S4355	B227; NIBSC 2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningiti	dis	A	5	ST-5 complex	5-1	9	F3-1
	9	, <u> </u>	Germany	1999	carrier	Neisseria meningiti	dis	В	930	ST-334 complex			
Ar		1 2 3 4	Germany	1999	carrier	Neisseria meningiti	dis	В	930	ST-334 complex			
5 (« Ar	, ) ()	1 2 3 4 🤇	Germany		carrier morphic sites Publication	-	dis	В	930	ST-334 complex			
5 (« Ar	9	1 2 3 4 tools	00	Poly		s Sequence bin	dis	В	930	ST-334 complex			
5 (« Ar	9 alysis 1 Breakdo	1 2 3 4 tools own: Fields Two Field ysis: BURST Codons	Combinations	Poly	morphic sites Publication	s Sequence bin		В	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 breakde	own of dataset			Help 🔽
Here you can create a table breaking Isolates	ng 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 a dataset orow 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

o field breakdown of dataset         eakdown of clonal_complex (MLST) by capsule_group:         ted option: Display values only.         Xes	Values and percentages         No value       1       Values and percentages         No value       1       Values and percentages         ST-11 complex       1       1         Values and percentages         ST-12 complex       1       1         ST-132 complex       1       1         ST-132 complex       1       2       2         ST-132 complex       1       1         ST-246 complex       2       1       2         ST-246 complex       2       2       2       2       2       2         ST-246 complex       2       2       2 <th 2"2"2"2"2"2"2"2"2"2"2"2"2"2"2"2"2"2<="" colspa="2" th=""><th><ul> <li>Oceanisme &gt; Main</li> </ul></th><th>asia ann 🔉</th><th>Maissar</th><th>in inclator</th><th>&gt; Diusias</th><th>&gt; Tue field</th><th>breakdauur</th></th>	<th><ul> <li>Oceanisme &gt; Main</li> </ul></th> <th>asia ann 🔉</th> <th>Maissar</th> <th>in inclator</th> <th>&gt; Diusias</th> <th>&gt; Tue field</th> <th>breakdauur</th>	<ul> <li>Oceanisme &gt; Main</li> </ul>	asia ann 🔉	Maissar	in inclator	> Diusias	> Tue field	breakdauur
Part Adown of Clonal_complex (MLST) by capsule_group: tet options: Display values only. Xex	Part A complex of Clonal_complex (MLST) by capsule_group: tet options: Display values only: Xex Show	> Organisms > Neiss	seria spp. >	Neisser	la isolates	s > Plugins	> Iwo neto	breakdown	
akdown of clonal_complex (MLST) by capsule_group:         ted options: Display values only.         xxx       Show         verse       Values and percentages         verse       Values and percentages         st-1 complex       1       3       1       4       1       1       12         ST-14 complex       1       1       1       1       3       5       5       5         ST-14 complex       1       1       1       3       5       5       5       5         ST-14 complex       1       1       1       3       5       <	akdown of clonal_complex (MLST) by capsule_group:         ted options: Display values only.         xxx       Show         verse       Values and percentages         verse       Values and percentages         st-1 complex       1       3       1       4       1       1       12         ST-14 complex       1       1       1       1       3       1       4       1       1       12         ST-14 complex       1       1       1       2       3       5       5       5       5       5         ST-14 complex       1       1       1       3       3       5	c: 111			<b>C</b> 1				
ted options: Display values only. Xerse Snow Values and percentages Values	ted options: Display values only. Xerse Snow Values and percentages Values	o field bre	akdo	wn (	of da	taset			
ted options: Display values only. Xers Show Values and percentages Values a	ted options: Display values only. Xers Show Values and percentages Values a								
ted options: Display values only. Xers Snow Values and percentages Values a	ted options: Display values only. Xers Snow Values and percentages Values a								
ted options: Display values only. Xerse Show Values and percentages Values	ted options: Display values only. Xerse Show Values and percentages Values	akdown of ch		mploy		T) by c	anculo	troup	
Axes       Show         evers       Values and percentages         al_complex (MLST)       A $\bullet$ B $\bullet$ C $\bullet$ E $\bullet$ N $\bullet$ X $\bullet$ Y $\bullet$ Z $\bullet$ I total $\bullet$ Structure       1       5       1       4       1       1       12         ST-1 complex       13       1       4       1       1       12         ST-11 complex       13       1       4       1       1       12         ST-12 complex       5       1       2       3       5         ST-147 complex       1       1       1       3       5         ST-130 complex       2       2       2       2       2       2         ST-140 complex       1       1       1       3       5       5       5       5         ST-140 complex       2       2       2       2       2       2       2       2       2       2         ST-240 complex       1       1       1       1       1       1       1       1       1         ST-320 complex       1       1       1       1       1       1       1	Axes       Show         evers       Values and percentages         al_complex (MLST)       A $\bullet$ B $\bullet$ C $\bullet$ E $\bullet$ N $\bullet$ X $\bullet$ Y $\bullet$ Z $\bullet$ I total $\bullet$ Structure       1       5       1       4       1       1       12         ST-1 complex       13       1       4       1       1       12         ST-11 complex       13       1       4       1       1       12         ST-12 complex       5       1       2       3       5         ST-147 complex       1       1       1       3       5         ST-130 complex       2       2       2       2       2       2         ST-140 complex       1       1       1       3       5       5       5       5         ST-140 complex       2       2       2       2       2       2       2       2       2       2         ST-240 complex       1       1       1       1       1       1       1       1       1         ST-320 complex       1       1       1       1       1       1       1		liai_co	mpiex		51) Uy C	apsule_	group.	
Values and percentages         Va	Values and percentages         Values and percentages         Values and percentages         No value       1       3       1       1       4       1       1       12         ST-11 complex       13       1       4       1       1       12         ST-11 complex       5       I       I       2       3         ST-12 complex       5       I       I       2       3         ST-12 complex       5       I       I       2       3         ST-13 complex       I       1       2       3         ST-132 complex       I       I       2       3         ST-132 complex       I       I       I       I       1         ST-132 complex       I	cted options: Display	alues only.						
Velues and percentages         Complex (MLST) × A × B × C × E × NC × V × Z × Cnl + Total +         No value       1       3       1       1       1       12         ST-1 complex       13       1       4       1       1       12         ST-1 complex       5       1       4       1       1       12         ST-11 complex       5       1       2       3       3       1       1       2       3         ST-11 complex       5       1       2       3       3       1       1       2       3         ST-14 complex       1       1       2       3       3       1       1       2       3         ST-14 complex       2       1       2       3       3       3       3       3       3         ST-18 complex       3       2 <t></t>	Values and percentages         No value       1       Values and percentages         No value       1       Values and percentages         ST-11 complex       1       1         Values and percentages         ST-12 complex       1       1         ST-132 complex       1       1         ST-132 complex       1       2       2         ST-132 complex       1       1         ST-246 complex       2       1       2         ST-246 complex       2       2       2       2       2       2         ST-246 complex       2       2       2 <th 2"2"2"2"2"2"2"2"2"2"2"2"2"2"2"2"2"2<="" colspa="2" td=""><td>Axes Show</td><td></td><td></td><td></td><td></td><td></td><td></td></th>	<td>Axes Show</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	Axes Show						
Capsule_group       •         al_complex (MLST) •       A •       B •       C •       N •       X •       Y •       Z •       Cotal •         No value       1       3       1       1       4       1       1       12         ST-1 complex       13       1       1       4       1       1       12         ST-11 complex       15       1       1       4       2       3         ST-167 complex       5       1       1       1       3         ST-174 complex       1       1       1       3         ST-178 complex       1       1       1       1         ST-188 complex       2       2       2         ST-280 complex       2       2       2         ST-280 complex       2       1       1         ST-240 complex       1       1       1         ST-260 complex       2       1       1       1         ST-340 complex       1       1       1       1         ST-260 complex       2       1       1       1         ST-340 complex       1       1       1       1	Cajsule_group       •         sal_complex (MLST) •       A •       B •       C •       N G •       W •       X •       Y •       Z •       Int •       Total •         No value       1       3       1       1       4       1       1       12         ST-1 complex       13       -       -       -       13       5         ST-11 complex       1       1       2       3       5         ST-167 complex       5       -       -       -       5         ST-174 complex       1       1       2       3       5         ST-18 complex       1       1       1       -       3       5         ST-197 complex       -       2       2       5       5         ST-174 complex       1       1       1       5       5         ST-18 complex       2       2       2       5       5         ST-20 complex       2       2       2       5       5         ST-26 complex       1       1       1       1       5         ST-34 complex       1       1       1       5       5         ST-								
All complex (MLST)       A       B       C       E       NG       W       X       Y       Z       Intel       Total         No value       1       3       1       1       1       1       12         ST-1 complex       13       1       4       1       1       12         ST-1 complex       13       1       4       1       1       12         ST-1 complex       13       1       4       1       1       12         ST-1 complex       5       1       1       1       17         ST-132 complex       1       1       1       1       5         ST-147 complex       1       1       1       1       5         ST-178 complex       1       1       1       1       5         ST-18 complex       5       2       2       2       2         ST-22 complex       1       1       1       1       1         ST-240 complex       2       2       1       1       1         ST-240 complex       2       1       1       1       1         ST-340 complex       6       1       1	Nat_complex (MLST)     A     B     C     E     NG     W     X     Y     Z     Ch     I       No value     1     3     1     1     4     1     1     12       ST-1 complex     13     1     4     1     1     13       ST-1 complex     17     5     6     17     15       ST-162 complex     5     6     1     2     5       ST-174 complex     1     1     1     1     5       ST-182 complex     1     1     1     1     1       ST-182 complex     1     1     1     1     1       ST-192 complex     1     2     2     2     2       ST-192 complex     1     1     1     1     1       ST-192 complex     1     1     1     1     1       ST-202 complex     1     2     2     2     2       ST-242 complex     1     1     1     1     1       ST-242 complex     1     1     1     1       ST-242 complex     1     1     1     1       ST-342 complex     1     1     1     1       ST-342 complex     1	everse Values a	ind percent	ages					
No value         1         3         1         1         4         1         1         12           ST-1 complex         13         -         -         -         13         -         13           ST-11 complex         5         -         -         -         17         -         5           ST-167 complex         5         -         1         2         3         -         5           ST-176 complex         -         1         1         1         -         3         -           ST-178 complex         -         1         1         1         -         3         -         5           ST-180 complex         -         1         1         1         -         -         1         -         5           ST-198 complex         -         2         2         -         2         2         5         -         5         5         -         1         1         5         5         5         5         5         5         1         1         1         5         5         5         5         5         5         5         5         5         5         5         5 <td>No value         1         3         1         4         1         1         12           ST-1 complex         13         -         -         -         -         13         -         13           ST-11 complex         5         -         17         -         2         33         -         -         -         -         -         -         -         17           ST-162 complex         5         -         1         1         2         3         -         1         -         1         -         1         -         1         -         1         -         1         1         1         1         -         1         &lt;</td> <td></td> <td></td> <td></td> <td>capsu</td> <td>ıle_group</td> <td></td> <td>¢</td>	No value         1         3         1         4         1         1         12           ST-1 complex         13         -         -         -         -         13         -         13           ST-11 complex         5         -         17         -         2         33         -         -         -         -         -         -         -         17           ST-162 complex         5         -         1         1         2         3         -         1         -         1         -         1         -         1         -         1         -         1         1         1         1         -         1         <				capsu	ıle_group		¢	
No value         1         3         1         1         4         1         1         12           ST-1 complex         13         -         -         -         13         -         13           ST-11 complex         5         -         -         -         17         -         5           ST-167 complex         5         -         1         2         3         -         5           ST-176 complex         -         1         1         1         -         3         -           ST-178 complex         -         1         1         1         -         3         -         5           ST-180 complex         -         1         1         1         -         -         1         -         5           ST-198 complex         -         2         2         -         2         2         5         -         5         5         -         1         1         5         5         5         5         5         5         1         1         1         5         5         5         5         5         5         5         5         5         5         5         5 <td>No value         1         3         1         4         1         1         12           ST-1 complex         13         -         -         -         -         13         -         13           ST-11 complex         5         -         17         -         2         33         -         -         -         -         -         -         -         17           ST-162 complex         5         -         1         1         2         3         -         1         -         1         -         1         -         1         -         1         -         1         1         1         1         -         1         &lt;</td> <td>al complex (MLST) +</td> <td>A + B +</td> <td>C + E +</td> <td>NG + \</td> <td>v + X + Y</td> <td>♦ Z ♦ cnl</td> <td>¢ Total ≎</td>	No value         1         3         1         4         1         1         12           ST-1 complex         13         -         -         -         -         13         -         13           ST-11 complex         5         -         17         -         2         33         -         -         -         -         -         -         -         17           ST-162 complex         5         -         1         1         2         3         -         1         -         1         -         1         -         1         -         1         -         1         1         1         1         -         1         <	al complex (MLST) +	A + B +	C + E +	NG + \	v + X + Y	♦ Z ♦ cnl	¢ Total ≎	
ST-1 complex       13       I       I       I       I         ST-11 complex       17       I       I       I7         ST-162 complex       5       I       I       I       S         ST-176 complex       I       I       I       I       S         ST-176 complex       I       I       I       I       S         ST-176 complex       I       I       I       I       S         ST-174 complex       I       I       I       I       I       S         ST-180 complex       I       I       I       I       I       S         ST-198 complex       I       I       I       I       I       S         ST-254 complex       I       I       I       I       I       S         ST-32 complex       I       I       I       I       S       S         ST-32 complex       I       I       I       I       S       S         ST-32 complex       I       I       I       I       S       S         ST-42 complex       I       I       I       I       S       S         ST-4 complex	ST-1 complex       13       I       I       I       I       I       I         ST-11 complex       5       IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII								
ST-162 complex     5     V     V     V     5       ST-167 complex     1     1     1     3       ST-174 complex     1     1     1     3       ST-174 complex     1     1     1     1       ST-178 complex     1     1     1     1       ST-178 complex     2     2     3       ST-178 complex     2     2     1       ST-180 complex     2     2     3       ST-220 complex     2     2     2       ST-220 complex     1     2     3       ST-230 complex     1     2     2       ST-250 complex     1     4     1       ST-250 complex     1     4     1       ST-250 complex     1     4     1       ST-350 complex     1     4     1       ST-350 complex     1     4     3       ST-41/44 complex     5     1     4       ST-450 complex     1     4     1       ST-450 complex     1     4     4       ST-450 complex     1     4     4       ST-50 complex     1     4     1       ST-50 complex     1     4     1       ST-	ST-162 complex     5     I     I     I     I     S       ST-147 complex     1     1     1     I     S       ST-174 complex     1     1     1     I     S       ST-174 complex     I     1     I     I     I       ST-174 complex     I     I     I     I     I       ST-174 complex     I     I     I     I     I       ST-178 complex     I     I     I     I     I       ST-198 complex     I     I     I     I     I       ST-220 complex     I     I     I     I     I       ST-220 complex     I     I     I     I     I       ST-230 complex     I     I     I     I     I       ST-254 complex     I     I     I     I     I       ST-254 complex     I     I     I     I     I       ST-340 complex     I     I     I     I     I       ST-354 complex     I     I     I     I     I       ST-340 complex     I     I     I     I     I       ST-41/44 complex     I     I     I     I       ST-50 complex								
ST-147 complex       1       1       1       1       1       1       3         ST-174 complex       1       1       1       1       1       3         ST-187 complex       3       3       3       3         ST-180 complex       3       2       4       3         ST-198 complex       3       2       2       2         ST-254 complex       1       1       1       1         ST-254 complex       1       1       1       1         ST-256 complex       1       1       1       1         ST-326 complex       1       1       1       1         ST-326 complex       1       1       1       1         ST-326 complex       1       1       1       1         ST-426 complex       1       1       1       1         ST-45 complex       1       1	ST-167 complex       1       1       1       2       3         ST-174 complex       1       1       1       3         ST-178 complex       3       1       4       3         ST-186 complex       3       2       4       3         ST-198 complex       3       2       4       5         ST-296 complex       2       2       4       2         ST-256 complex       1       4       1       1         ST-326 complex       1       4       4       1         ST-326 complex       1       4       4       1         ST-326 complex       1       4       4       1         ST-40 complex       1       4       4       1         ST-40 comp			17					
ST-174 complex       1       1       1       1       3         ST-178 complex       3       1       1       1         ST-180 complex       3       2       2       2         ST-198 complex       3       2       2       2         ST-226 complex       1       2       2       2         ST-236 complex       1       1       1         ST-256 complex       1       1       1         ST-256 complex       1       1       1         ST-256 complex       1       1       1         ST-353 complex       1       1       1         ST-354 complex       1       1       1         ST-352 complex       1       1       1         ST-353 complex       1       1       1         ST-353 complex       1       1       1         ST-40 complex       1       1       1         ST-41/44 complex       5       1       1         ST-50 complex       1       1       1         ST-50 complex       1       1       2         ST-50 complex       1       1       1         ST-50 complex </td <td>ST-174 complex       1       1       1       1       3         ST-174 complex       1       1       1       1       3         ST-18 complex       3       1       1       1       1         ST-198 complex       3       2       2       2       2         ST-206 complex       1       2       2       2       2         ST-256 complex       1       1       1       1         ST-353 complex       1       1       1       1         ST-354 complex       1       1       1       1         ST-354 complex       1       1       1       1         ST-354 complex       1       1       1       1         ST-41/44 complex       5       1       1       1         ST-50 complex       1       1       1       1         ST-50 complex       1       1       1       2         ST-41/44 complex       5       1</td> <td></td> <td>5</td> <td></td> <td></td> <td></td> <td>2</td> <td>5</td>	ST-174 complex       1       1       1       1       3         ST-174 complex       1       1       1       1       3         ST-18 complex       3       1       1       1       1         ST-198 complex       3       2       2       2       2         ST-206 complex       1       2       2       2       2         ST-256 complex       1       1       1       1         ST-353 complex       1       1       1       1         ST-354 complex       1       1       1       1         ST-354 complex       1       1       1       1         ST-354 complex       1       1       1       1         ST-41/44 complex       5       1       1       1         ST-50 complex       1       1       1       1         ST-50 complex       1       1       1       2         ST-41/44 complex       5       1		5				2	5	
ST-178 complex       1       1       1         ST-180 complex       3       2       3         ST-198 complex       2       2       2         ST-22 complex       2       2       2         ST-22 complex       1       1       1         ST-22 complex       1       1       1         ST-22 complex       1       1       1         ST-25 complex       1       1       1         ST-25 complex       1       1       1         ST-25 complex       1       1       1         ST-32 complex       1       1       1         ST-35 complex       1       1       1         ST-35 complex       1       1       1         ST-4 complex       1       1       1         ST-4 complex       1       1       1         ST-5 complex       1       1       1         ST-4 complex       1       1       1         ST-5 complex       1       1       1         ST-5 complex       1       1       1         ST-50 complex       1       1       1         ST-60 complex       1	ST-178 complex       3       1       1       1         ST-180 complex       2       2       3         ST-198 complex       2       2       2         ST-22 complex       2       2       2         ST-22 complex       1       2       2         ST-22 complex       1       1       1         ST-22 complex       1       1       1         ST-25 complex       1       1       1         ST-25 complex       1       1       1         ST-26 complex       1       1       1         ST-25 complex       1       1       1         ST-35 complex       1       1       5         ST-35 complex       1       1       1         ST-46 complex       1       1       1         ST-50 complex       1       1       1         ST-50 complex </td <td></td> <td></td> <td>1</td> <td></td> <td></td> <td>2</td> <td></td>			1			2		
ST-18 complex     3     2     3       ST-19 complex     2     2       ST-29 complex     1     2       ST-25 complex     1     1       ST-35 complex     1     1       ST-35 complex     1     1       ST-35 complex     1     4       ST-35 complex     1     4       ST-35 complex     1     4       ST-45 complex     1     5       ST-4 complex     1     6       ST-55 complex     1     4       ST-4 complex     1     6       ST-55 complex     1     4       ST-55 complex     1     6       ST-55 complex     1     4       ST-55 complex     1     6       ST-55 complex     1     1       ST-50 complex     1     1       ST-50 complex     1     9	ST-18 complex     3     3     3       ST-198 complex     2     2     2       ST-296 complex     1     2     2       ST-254 complex     1     1     1       ST-354 complex     1     1     1       ST-45 complex     1     1     1       ST-46 complex     1     1     1       ST-53 complex     1     1     1       ST-40 complex     1     1     1       ST-40 complex     1     1     1       ST-53 complex     1     1     1       ST-50 complex     1     1     9       Iotal     2     1     9			1					
ST-198 complex       2       2       2         ST-22 complex       1       2       2         ST-32 complex       1       1       1         ST-35 complex       1       1       1         ST-26 complex       2       1       1         ST-35 complex       1       1       1         ST-4(Ard complex       1       1       1         ST-4(Ard complex       1       1       1         ST-52 complex       1       1       1         ST-4(Ard complex       1       1       2         ST-52 complex       1       1       2         ST-50 complex       1       1       2         ST-50 complex       1       1       2         ST-50 complex	ST-198 complex       2       2       2         ST-220 complex       1       2       2         ST-23 complex       1       1       1         ST-23 complex       1       1       1         ST-23 complex       1       1       1         ST-25 complex       1       1       1         ST-25 complex       2       1       1         ST-25 complex       2       1       1         ST-25 complex       2       1       1         ST-35 complex       2       1       5         ST-35 complex       1       5       6         ST-35 complex       1       5       1       6         ST-41/44 complex       5       1       6         ST-50 complex       1       6       5         ST-41/44 complex       5       1       6         ST-50 complex       1       1       1         ST-50 complex       1       1       2         ST-50 complex       1       1       9         Total       28       27       18       1       3       1         Total       28       27       18 <td></td> <td>3</td> <td></td> <td></td> <td></td> <td></td> <td>3</td>		3					3	
ST-226 complex       1       1         ST-254 complex       1       1         ST-254 complex       1       1         ST-254 complex       1       1         ST-254 complex       1       1         ST-326 complex       2       1       1         ST-326 complex       2       1       3         ST-354 complex       1       1       1         ST-45 complex       1       1       1         ST-4 complex       5       1       6         ST-45 complex       1       1       1         ST-50 complex       1       1       1         Totat       28       27       18       1       3       1         ST-50 complex       1       1	ST-226 complex       1       1       1         ST-254 complex       2       1       4       3         ST-326 complex       6       6       5       3         ST-354 complex       1       6       1       5         ST-354 complex       1       6       1       1         ST-354 complex       1       6       1       1         ST-45 complex       1       6       1       5         ST-4 complex       5       1       6       6         ST-4 complex       1       6       1       6         ST-5 complex       1       6       6         ST-5 complex       1       1       2         ST-6 complex       1       1       2         ST-35 complex       1       1       9         Total       2       2       1       3       1       99	ST-198 complex							
ST-23 complex       1       1       1         ST-254 complex       2       1       1         ST-252 complex       2       1       3         ST-32 complex       6       6       6         ST-33 complex       1       6       1         ST-35 complex       1       6       1         ST-47 complex       1       6       1         ST-47 complex       1       6       1         ST-44/44 complex       5       1       6         ST-45 complex       1       6       6         ST-46 complex       1       6       6         ST-41/44 complex       5       1       6         ST-55 complex       1       6       1         ST-56 complex       1       1       2         ST-50 complex       1       1       2         ST-50 complex       1       1       99	ST-23 complex       I       I       1         ST-254 complex       I       I       I         ST-254 complex       I       I       I         ST-35 complex       I       I       I         ST-40 complex       I       I       I         ST-41/44 complex       I       I       I         ST-50 complex       I       I       I       I         ST-41/44 complex       I       I       I       I         ST-50 complex       I       I       I       I         ST-50 complex       I       I       I       I         ST-50 complex       I       I       I       I         Total       I       I       I       I       I         Total       I       I       I       I       I					2			
ST-254 complex       1       1       1         ST-256 complex       2       1       3         ST-326 complex       1       6       6         ST-336 complex       1       1       1         ST-337 complex       1       1       1         ST-357 complex       1       1       1         ST-45 complex       1       1       1         ST-45 complex       5       1       6         ST-44 complex       5       1       6         ST-44 complex       5       1       6         ST-45 complex       1       1       6         ST-46 complex       1       1       1         ST-50 complex       1       1       1         ST-50 complex       1       1       2         ST-60 complex       1       1       2         ST-60 complex       1       1       99         Total       28       27       18       3       1       1         St-60 complex       1       3       1       1       99	ST-254 complex       1       1       1         ST-254 complex       2       1       3         ST-320 complex       1       6       6         ST-334 complex       1       1       1         ST-353 complex       1       1       1         ST-35 complex       1       1       1         ST-45 complex       1       1       1         ST-45 complex       1       6       8         ST-44 complex       5       1       6         ST-44 complex       5       1       6         ST-461 complex       1       1         ST-55 complex       1       1       2         ST-56 complex       1       1       2         ST-60 complex       1       1       2         ST-60 complex       1       1       9         Total       28       27       18       3       1       1         Iorat       28       27       18       1       3       1       9		1				1		
ST-262 complex       2       1       3         ST-32 complex       6       6         ST-32 complex       1       6         ST-35 complex       1       1         ST-35 complex       1       1         ST-4 complex       1       1         ST-4 complex       5       1       6         ST-4 complex       1       1       1         ST-4 complex       1       6         ST-4 complex       1       6         ST-5 complex       1       6         ST-50 complex       1       1       2         ST-60 complex       1       1       2         ST-60 complex       1       1       9         Totat       28       27       18       5       1       3         ST-60 complex       1       99       1       1       1	ST-269 complex       2       1       3         ST-324 complex       6       6         ST-354 complex       1       1         ST-355 complex       1       1         ST-4 complex       1       1         ST-4 complex       1       1         ST-4 complex       5       1       6         ST-4 complex       5       1       6         ST-4 complex       5       1       6         ST-4 complex       1       1       1         ST-4 complex       1       6         ST-4 complex       1       6         ST-5 complex       1       6         ST-5 complex       1       1       2         ST-6 complex       1       1       2         ST-6 complex       1       3       1       99				1		1		
ST-32 complex       6         ST-334 complex       1         ST-354 complex       1         ST-35 complex       1         ST-41 complex       5         ST-41 complex       1         ST-55 complex       1         ST-55 complex       1         ST-55 complex       1         ST-55 complex       1         ST-56 complex       1         ST-50 complex       1         ST-50 complex       1         ST-60 complex       1 <t< td=""><td>ST-32 complex       6         ST-334 complex       1         ST-353 complex       1         T-45 complex       1         ST-441 complex       5         ST-441 complex       5         ST-450 complex       6         ST-461 complex       1         ST-55 complex       6         ST-55 complex       1         ST-50 complex       1         ST-60 complex       1</td><td></td><td>2</td><td></td><td></td><td></td><td></td><td></td></t<>	ST-32 complex       6         ST-334 complex       1         ST-353 complex       1         T-45 complex       1         ST-441 complex       5         ST-441 complex       5         ST-450 complex       6         ST-461 complex       1         ST-55 complex       6         ST-55 complex       1         ST-50 complex       1         ST-60 complex       1		2						
ST-35 complex     1     1       ST-4 complex     5     1     8       T-41/44 complex     5     1     6       ST-45 complex     1     4     1       ST-53 complex     1     4     1       ST-55 complex     1     1     2       ST-50 complex     1     1     2       ST-50 complex     1     1     99	ST-35 complex       1       1       1         ST-4 complex       8       8       8         ST-41/44 complex       5       1       6         ST-45 complex       1       1       6         ST-46 complex       1       1       6         ST-50 complex       1       1       1         ST-50 complex       1       1       2         ST-60 complex       1       1       99								
ST-4 complex     8     8       ST-41/44 complex     5     1     6       ST-41/44 complex     1     6       ST-61 complex     1     6       ST-55 complex     6     6       ST-55 complex     1     1       ST-60 complex     1     1       Totat     28     27     18     3     14     3     1     1	ST-4 complex     8     8       ST-41/44 complex     5     1     6       ST-41/44 complex     1     6       ST-50 complex     6     6       ST-53 complex     1     1       ST-50 complex     1     1       ST-60 complex     1     1       ST-60 complex     1     1       ST-60 complex     1     1       ST-60 complex     1     99		1						
5T-41/44 complex     5     1     6       5T-461 complex     1     1       5T-5 complex     6     6       ST-50 complex     1     1       ST-60 complex     1     1       Totat     28     27     18     5     1	ST-41/44 complex     5     1     6       ST-461 complex     1     1       ST-56 complex     6     6       ST-53 complex     1     1       ST-60 complex     1     1       Torat     28     27     18     3     1     1				1				
ST-461 complex       1       1       1       1         ST-5 complex       6       1       6         ST-53 complex       1       1       2         ST-60 complex       1       1       2         Total       28       27       18       3       1       3       1       99	ST-461 complex       1       2       2       1         ST-5 complex       6       2       6       6         ST-53 complex       1       1       2         ST-60 complex       1       2       1         Total       28       27       18       3       1       3       1       99				1				
ST-5 complex     6     6       ST-53 complex     1     1       ST-50 complex     1     1       Total     28     27     18     3     1     3     1	ST-5 complex       6       6         ST-53 complex       1       1       2         ST-60 complex       1       1       2         Total       28       27       18       3       1       3       1       99				-				
ST-60 complex         1         1         1           Total         28         27         18         3         1         3         1         99	ST-60 complex         1         1         1           Total         28         27         18         3         14         3         1         3         1         99	ST-5 complex						6	
Total 28 27 18 3 14 3 1 3 1 99	Total 28 27 18 3 14 3 1 3 1 99				1		1		
			20 27				7 4 4		
		Iotal	28 27	18 5	14	5 1	5 1 1	99	
		ĭ X							
arts									
		to enlarge							
	to enlarge								
	to enlarge	Value	S		100% -	P	ercentages		
to enlarge	Values Dercentanes	28 -			90% -				
to enlarge Values Percentages 99% - Percentages 99% - Percentages 100% - Percentages 100	Values Percentages 100% → Perce	24 -			70% -				
to enlarge	Values Percentages								
to enlarge	Values Va	16 -			40% -				
to enlarge	Values Values Percentages 90% 90% 90% 90% 90% 90% 90% 90%	18 - 16 - 14 - 12 -							
to enlarge	Values Percentages								
to enlarge	Values Va		<b></b>		10% -				
to enlarge	Values Va		+	l	10% -	P C	40 +	1 1	

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 field 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		
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.	🗠 ANALYSIS	
			+
			+
		E TYPING	

Jump to the 'Breakdown' category, follow the link to Unique Combinations, then click 'Launch Unique Combinations'.

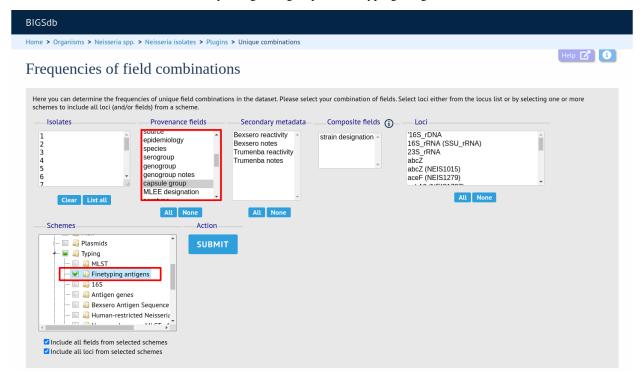
The function can be selected by clicking the 'Unique combinations' link in the Breakdown section of the main contents page. This will run the analysis on the entire database.

Unique Combinations
Summary: Determine frequencies of unique field combinations
The Unique Combinations plugin calculates the frequencies of unique file combinations within an isolate dataset. Primary metadata fields, allele designations and scheme fields can be combined. Results are returned in an Excel spreadsheet.
Documentation bigsdb.readthedocsio 🗗
ACID         Terpencies of field combinations         Image: Combination of the states are st

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	TDAIM	RAA! INIROC 7140: 5104A	Philippines	1209		iveisseria meningiu	ais	A 1	SI-1 complex	<b>D-</b> 7	10	F2-T
	0120/95		Czech Republic		invasive (unspecified/other)	Neisseria meningiti		X 117			14	
.5	1		Germany	1999	carrier	Neisseria meningiti		E 864				
16	2		Germany	1999	carrier	Neisseria meningiti		B 854	ST-18 complex			
L7	3		Germany	1999	carrier	Neisseria meningiti		W 174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningiti		B 19	ST-18 complex	_		
.9	S3131	B213; NIBSC_2813; Z1213			invasive (unspecified/other)	Neisseria meningiti		A 4	ST-4 complex	7	13-1	F1-5
20	5		Germany	1999	carrier	Neisseria meningiti		NG 198	ST-198 complex			
1	6		Germany	1999	carrier	Neisseria meningiti			ST-198 complex			
22	7		Germany	1999	carrier	Neisseria meningiti		E 60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningiti		B 32	ST-32 complex			
24	S4355	B227; NIBSC_2806; Z1227			invasive (unspecified/other)	Neisseria meningiti		A 5	ST-5 complex	5-1	9	F3-1
	9		Germany	1999	carrier	Neisseria meningiti	dis	B 930	ST-334 complex			
, (	00	1 2 3 4 5 6	ക	<u> </u>								
	alysis 1	tools		Poly	morphic sites Publication	Sequence hin						
(« Aı	alysis f	tools own: Fields Two Field	Combinations		morphic sites Publication							
(« Ar	alysis 1	tools own: Fields Two Field	Combinations Gene Presence		morphic sites Publication me Comparator BLAST		PCR					
(« Ar	alysis f	tools own: Fields Two Field yysis: 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:         fnished           Start time:         2020-08-01 09:03:26           Progress:         100%           Stop time:         2020-08-01 09:03:28           Total time:         2 seconds
Output Number of unique combinations: 105 Files
Combinations table (text) Combinations table (Excel) Tar file containing all output files
Please note that job results will remain on the server for 7 days.

#### CHAPTER

# **FIFTEEN**

# **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									
Neisseria isolates database									
Query database	+3 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.	Export dataset Contigs Sequences								
	MALYSIS								
		+							
		+							
	S TYPING								

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	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
50	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
51	393	B392; NIBSC_2823; Z1392	Greece	1968	carrier	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
54	254	B439; NIBSC_2812; Z1439	Djibouti	1966	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F1-7
57	S5611	B466; NIBSC_2765; Z1466	Australia	1977	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex	5-2	10	F5-1
32	11-004	B503; NIBSC_2826; Z1503	China	1984	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex	20	9	F3-8
34	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

11141, 515 1001	0			
Rreakdown:	Fields	Two Field	Combinations	F

🕓 Breakdown:	Fields	Two Field	Combinatio	ons Polyme	orphic sites	Publication	s Sequence bin	
🛃 Analysis:	BURST	Codons	Gene Preser	ce Genom	e Comparato	BLAST	rMLST species id	PCR
Export:	Dataset	Contigs	Sequences					
🗹 Third party:	GrapeTre	e iTOL	PhyloViz	Microreact				

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 > Neisseria spp. > Neisseria isolates > Plugins > Contigs											
Contig analysis	Contig analysis and export										
	: ids from which contigs are associated - use Ctrl or Shift to make n agged within the contig - if these loci overlap then the total tagged										
Isolates	Options	Filter by									
1 2 7 10 11 13 ~	Identify contigs with >= 0	Sequence method: Project: Experiment: Minimum length: ()	<b>~</b> ()								
10 Clear List all	RESET SUBMIT										

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			•	Action		Minimum length:
	Clear	List all	[	RESET	SUBMIT	
Cor	ntigs wi	th >=0%	-		gth untagged	
id ≑	isolate 🕈	contias ÷	matching	contigs 🗢	non-matching contig	
iu •		contrigs +	count 🕈 d	lownload 🕈	count 🕈 download	
1	A4/M1027	364	364	*	0	
2	120M	359	359		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 64	393 254	282	282		0	
67	254 S5611	251 194	251 194		0	
82	11-004	258	258	-	0	
84	IAL2229	188	188	-	0	
90	CN100	236	236		0	
					, , , , , , , , , , , , , , , , , , ,	

You can also download all the data in a tar file by clicking the 'Batch download' link.

11 13			•			Minimum length:	
19			• 1e	Action			
	Clear	List all		RESET	SUE	міт	
6		.1		,			
Cor	itigs wi	th >=09	% sequ	ience leng	gth unta	78eg	
			matchir	ng contigs 🔹	non-matcl	ing contigs 🕈	
id \$	isolate 🕈	contigs ¢	count +	download \$	count +	download +	
1	A4/M1027	364	364	*	0		
2	120M	359	359	*	0		
7	7891	199	199	4444444444444444444444	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	IAL2229	188	188	*	0		
90	CN100	236	236	*	0		
E	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 🗸 % of sequence untagged	1
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	SUBN	іт		
Cor	ntigs wi	th >=50			ngth unta			
id 🕈	isolate 🕈	contigs +						
		744	count + d			ownload 🕈		
	A4/M1027 120M	364 359	141 60	*	223 299	4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4		
2	7891	199	38	****	161			
10	6748	652	377	-	275			
11	129E	272	55		217	1		
13	139M	293	164	*	129	*		
19	S3131	173	35	*	138	*		
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	*		
52	243	229	45	*	184	*		
61	393	282	49		233	*		
64	254	251	47	*	204	*		
67	S5611	194	42	-	152	*		
82	11-004	258	37	-	221			
84 90	IAL2229 CN100	188	30 46		158 190	<u>×</u>		
90	CINIOO	236	40	-	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	

### *Neisseria* isolates database

Query database	+) LOG IN	
Search database		SIONS
Browse, search by any criteria, or enter list of attributes.		rs +
	C EXPORT	-
Search by combinations of loci This can include partial matches to find related isolates.	Export dataset Contigs Sequences	]
		S
	CUSTOM	IISE +
		ATION +

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.

16	2			Germany	1999	carrie	er	Neisseria meningi	tidis	В	854	ST-18 complex			
7	3			Germany		carrie	er	Neisseria meningi	tidis	W	174	ST-174 complex			
8	4			Germany		carrie		Neisseria meningi		В	19	ST-18 complex			
9	S3131	B213; NIBSC_2	2813; Z1213			invasive (unspe		Neisseria meningi		A	4	ST-4 complex	7	13-1	F1-5
0	5			Germany		carrie		Neisseria meningi		NG	198	ST-198 complex			
1	6			Germany		carrie		Neisseria meningi		NG	198	ST-198 complex			
2	7			Germany		carrie		Neisseria meningi		E	60	ST-60 complex			
3	8			Germany		carrie		Neisseria meningi		В	32	ST-32 complex			
4	S4355	B227; NIBSC_2	2806; Z1227			invasive (unspe		Neisseria meningi		A	5	ST-5 complex	5-1	9	F3-1
5	9			Germany	1999	carrie	er	Neisseria meningi	tidis	В	930	ST-334 complex			
@	)()	123	4	$\odot$											
« Ar	alysis 1		4	00											
	alysis f	tools	4	Combinatio	ns Poly	morphic sites	Publications	Sequence bin							
		tools own: Fields	4 Two Field Codons	Combinatio Gene Present		morphic sites ome Comparator		Sequence bin rMLST species id	PCR						
	Breakdo	tools own: Fields lysis: BURST	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 References
1       id       isolate       is
Schemes Classification schemes Options
Image: Second
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.

Molecular weights

Export protein molecular weights GTG/TTG at start codes for methionine

Click the 'Export protein molecular weight' checkbox. Additional columns (or rows depending on the output format) will be created to include the molecular weight data.

# **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>B</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.

**		1.		'	~	**		ST 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	
(	«)(	$\checkmark$		12	3	4	5 (	• <b>• •</b> •

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	lect	
This script will export allelic profiles in tab-delimite Select STs	d text and Excel formats.	
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 Include datestamps 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           Progress:         100%           Stat time:         220-08-01 14:27:33           Total time:         12 seconds
Output         Files         Image: Profiles - Tab-delimited text (text)         Profiles - Tab-delimited text         Image: Profiles -
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		
Home > Organisms > Neisseria spp. > Neisseria isolates		

Neisseria isolates database

Query database		DI LOG IN	
Search database			
Browse, search by any criteria, or enter list of attributes.	I I	PROJECTS	+
	C	Z EXPORT	-
Search by combinations of loci This can include partial matches to find related isolates.		Export dataset Contigs Sequences	
	L	ANALYSIS	
	1		+
	•	INFORMATION	+
	ł	S TYPING	

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.

					×///							JI 17 I COMPLEX			
18	4			Germany	1999	carrier		eisseria meningit		В	19	ST-18 complex			
19	S3131	B213; NIBSC_2	2813; Z1213			invasive (unspecified		eisseria meningit		А	4	ST-4 complex	7	13-1	F1-5
20	5			Germany	1999	carrier	Ne	eisseria meningit	idis	NG	198	ST-198 complex			
21	6			Germany	1999	carrier	Ne	eisseria meningit	idis	NG	198	ST-198 complex			
22	7			Germany	1999	carrier	Ne	eisseria meningit	idis	E	60	ST-60 complex			
23	8			Germany	1999	carrier	Ne	eisseria meningit	idis	В	32	ST-32 complex			
24	S4355	B227; NIBSC_2	2806; Z1227	Denmark	1974	invasive (unspecified	d/other) Ne	eisseria meningit	idis	A	5	ST-5 complex	5-1	9	F3-1
25	9			Germany	1999	carrier	Ne	eisseria meningit	idis	В	930	ST-334 complex			
«	$\odot$	123	4	$\odot$											
« An	alysis f	1 2 3 tools	4	$\odot$											
	alysis 1		4	Combinations	s Poly	morphic sites Pub	blications	Sequence bin							
		own: Fields	4 Two Field Codons	Combinations Gene Presence		- · · · · · · · · · · · · · · · · · · ·		Sequence bin ILST species id	PCR						
	Breakdo	own: Fields lysis: BURST	Codons			- · · · · · · · · · · · · · · · · · · ·			PCR						

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.

files. Only DNA loci that have a corresponding datab	Multi-FASTA (XMFA) format suitable for loading into third-pa ase containing allele sequence identifiers, or DNA and peptid ne or more schemes to include all loci that are members of th	e loci with genome sequences tagged, can be incl	luded. Please check the loci
Aligned output is limited to 200 records; total output	It (records x loci) is limited to 1,000,000 sequences.		
Please be aware that if you select the alignment op	tion it may take a long time to generate the output file.		
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 date received non culture v	16S_rDNA 16S_rRNA (SSU_rRNA) 23S_rRNA abcZ abcZ (NEIS1015) aceF (NEIS1279) All None Paste list	
Schemes	Options	Action	
Plasmids     Typing     MLST     Finetyping antigens     165	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	SUBMIT	
🔲 🌆 Antigen genes 📃 🌆 Bexsero Antigen Sequence	✓ Do not include incomplete sequences Include 0 ✓ bp flanking sequence ()		
Human-restricted Neisseria	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 im:         2020-08-01 15:26:19           Total time:         6 seconds
Output Files
XMFA output file (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.

If both allele designations and tagged sequences exist for a locus, choose how you want these handled: (i)
0
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 v bp flanking sequence 🧃
✓ 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.

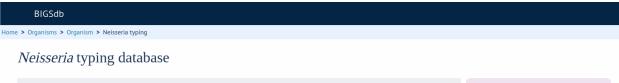
# CHAPTER

# SIXTEEN

# 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.



Query a sequence	Find alleles	Search for allelic profiles		⇒) LOG IN	
Sequence query	Sequence attribute search	Allelic profile query			
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	+
				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.			+
		Batch profile query			+
		Lookup multiple allelic profiles together.		SISOLATES	

# 16.1 Registering a user account

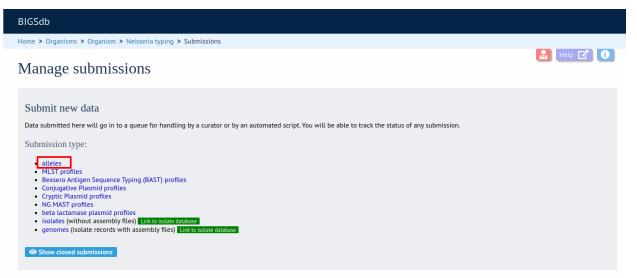
You must have an account for the appropriate database in order to use the submission system. On systems utilizing site-wide databases, such as PubMLST, this can be done automatically via the web. Other sites may require you to contact a curator to set this up.

# 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.

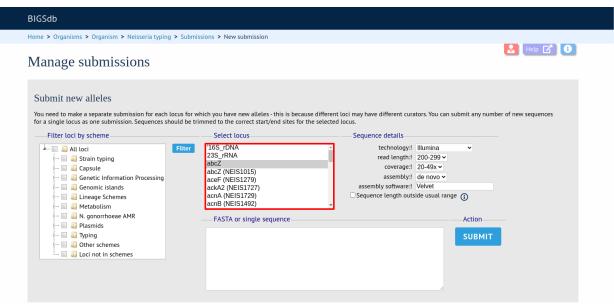
# 16.2.1 Start

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:



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 > Submissi	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			ubmit any number of new sequences
	Select locus	Sequence details	
Alt loci     Genetic Information Processing     Genemic Islands     Genemic Islan	Filter 165_rDNA 235_rRNA abcZ abcZ (NEIS1015) accF4 (NEIS1279) acrA2 (NEIS1727) acrA2 (NEIS1727) acrA3 (NEIS1729) acrB3 (NEIS1492) 	technology: read length: coverage: assembly: assembly: Sequence length outsi	200-299 ▼ 20-49x ▼ de novo ▼ Velvet
Genomic Islands     G	ackA2 (NEIS1727) acnA (NEIS1729) acnB (NEIS1492)	assembly software:	Velvet de usual range 🕦

The locus list is now constrained making selection easier.

BIGSdb		
Home > Organisms > Organism > Neisseria typing > Submissions > Ne	> New submission	
Manage submissions	🔒 Help 🗹 🚺	
Submit new alleles You need to make a separate submission for each locus for which you have for a single locus as one submission. Sequences should be trimmed to the Filter loci by scheme Genetic Information Processing Genetic Information Proceson Genetic Information Processing G	Select locusSequence details	
🛄 🤐 NG MAST		

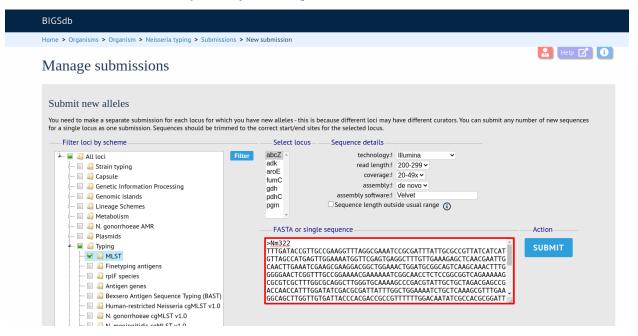
### 16.2.3 Enter details of sequencing method

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.

Home > Organisms > Organism > Neisseria typing > Su	Dmissions > New Submission		
Manage submissions			💄 Help 🕜 🤃
Error: Sequence "Nm324" has already been defined Submit new alleles You need to make a separate submission for each locus f		erent loci may have different curators. You can submit ar	y number of new sequences
for a single locus as one submission. Sequences should b Filter loci by scheme	e trimmed to the correct start/end sites for the sele Select locus	ected locus. Sequence details	
All loci All loci All loci Capsule Capsule Capsule Capsule	r '16S_rDNA 23S_rRNA abcZ (NEIS1015) ackF2 (NEIS127) ackF2 (NEIS1727)	technology: Illumina ▼ read length: 100-199 ▼ coverage: 20-49× ↓ assembly: de novo ▼ assembly software: Spades	
	acnA (NEIS1729) acnB (NEIS1492)	Sequence length outside usual range (j	

Assuming the preliminary checks have passed you will then be able to add additional information to your submission.

me > Organisms > Organism > Neisseria	a typing > Submissions > New submission	
lanage submissions		Help 🔽 🕄
Supporting files Please note that if you are submitting allele ack or compress your files (zip, rar etc).	es determined by Sanger sequencing that you must upload both forward	and reverse trace files. Your submission will be rejected if you don't. Please do not
Abort submission Submission: BIGSdb_20200	802122845_002379_40069	
	u can upload up to 64 MB in one go (the upload may fail if you try to do i	xplanatory note so that they can be linked to the appropriate submission item. more than this - just try again with fewer files at a time if it does), although you
	Drop files here or click to	o upload.
Sequences	Drop files here or click to	Sequence details
You are submitting the following abcZ see Identifier Length Nm322 433 TTTGATACCGTTGCCC		Sequence details technology: Illiumina v read length: 200-299 v

# 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.

E-mail
Updates will be sent to keith.jolley@zoo.ox.ac.uk.
E-mail submission updates
Messages
Thanks!
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.

E-mail		
Updates will be sent to keith.jolley@zoo.ox.ac.uk.		
E-mail submission upda Messages	ates	
Timestamp	User	Message
2020-08-02 11:30:58+00	Keith Jolley	Thanks!
	Message:	Append
	message.	Appenu

# 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.

Supporting files Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 64 MB. You can upload 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 can upload multiple times so that the total size of the submission can be larger.				
Drop files here or click to u	pload.			
Sequences	Sequence details			
You are submitting the following abcZ sequences: Download res Identifier Length Sequence Status Query Assigned allele Nm322 433 TTTGATACCGTTGCCGAAGG GCGGATTGTCGAACTCGATC pending Q Nm323 433 TTTGATACTGTTGCCGAAGG GCGGATTGTCGAACTCGATC pending Q	technology: Illumina read length: 200-299 coverage: 20-49x assembly: de novo assembly: Uvlvet Sequence length outside usual range ()			

The files will be uploaded and shown in a table.

Supporting files Please upload any supporting files required fo	or curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item.
Individual filesize is limited to 64 MB. You can can upload multiple times so that the total si	n upload 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 ze of the submission can be larger.
	Drop files here or click to upload.
Uploaded files Filename Size Delete NM322_abcZ_Rab1 252 KB NM322_abcZ_Fab1 269.6 KB NM322_abcZ_Fab1 281.8 KB NM322_abcZ_Rab1 281.8 KB Delete selected files	

Files can be removed from the submission by checking the appropriate 'Delete' box and clicking 'Delete selected files'.

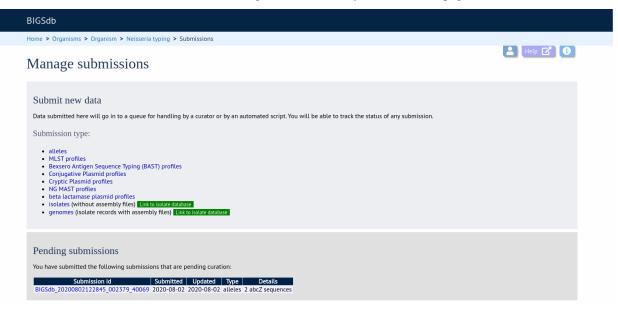
# 16.2.7 Finalize submission

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
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	

Your submission will then be listed under 'Pending submissions' on your submission page.

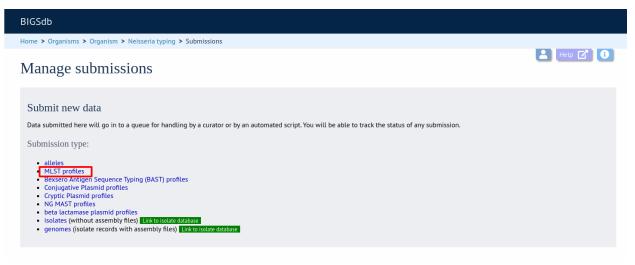


# 16.3 Profile submission

### 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 > Neisse	eria typing > Submissions > New s	submission		
Manage submission	S			🛓 Help 🔽 🚯
Submit new MLST profiles Paste in your profiles for assignment usin Templates Please paste in tab-delimited ter id <u>abcZ adk aro</u> 5 7 3 3 6 4 9 6 9	ng the template available below. xt (include a field header line) DE fumC gdh pdhC	2 <u>pgm</u> 32 2 6 9	Action	

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.					
Paste in y	t new M our profiles f	-		ne template	available I	oelow.				
Templ	X	tab dalim	sited tout (	nelude a fi	old boods	r line)			Artion	
id	abcZ abcZ 8 5 3 9	adk 5 7 6 6	aroE 32 3 4 9	fumC 2 6 3 9	gdh 6 33 8 9	pdhC 43 12 4 6	pgm 32 2 6 9		Action	

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		Help 🕜 🚯
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 nam Individual filesize is limited to 64 MB. You can upload up to 64 MB in one go (the t 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     Control of the second secon	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.

Timestamp	User	Message
020-08-02 15:40:36+00	Keith Jolley	Corresponding isolate submission has been made. BIGSdb_20200802163741_003123_4642

# 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!'.

E-mail		Action
Updates will be sent to keith.jo	lley@zoo.ox.ac.uk.	FINALIZE SUBMISSION!
E-mail submission updates		TIRALIZE SUBMISSION.

Your submission will then be listed under 'Pending submissions' on your submission page.

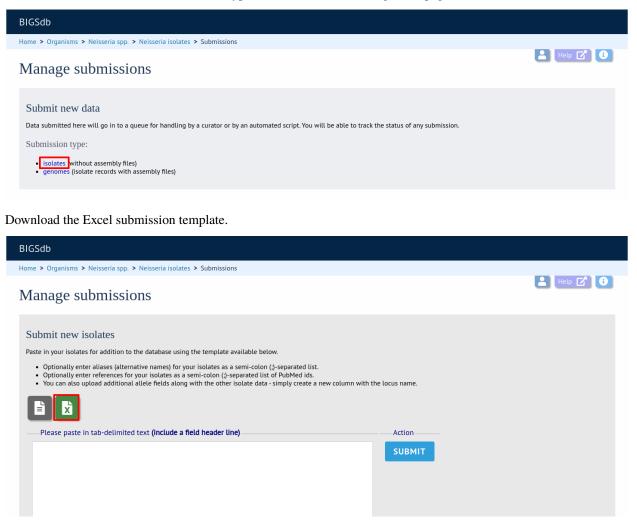
BIGSdb	
Home > Organisms > Organism > Neisseria typing > Submissions	
Manage submissions	Letp 📝 🚯
Submit new data Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission. Submission type:      alleles     MLST profiles     Mexsero Antigen Sequence Typing (BAST) profiles     Conjugative Plasmid profiles     Cryptic Plasmid profiles     Not MAST profiles     beta lactamase plasmid profiles     isolates (without assembly files) Link to isolate database     genomes (isolate records with assembly files) Link to isolate database	
Submissions         Submission id       Submitted Updated Type Details         BIGSdb_20200802163741_003907_56760       2020-08-02       profiles       3 MLST profiles	

# 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.



# 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.

SSdb						
ne > Organisms > Neisseria spp. > Neis	ria isolates > Submissions					
lanage submissions		Help 🗹 🕄				
ubmit new isolates						
aste in your isolates for addition to the da	base using the template available below.					
· Optionally enter references for your i	<ul> <li>Optionally enter aliases (alternative names) for your isolates as a semi-colon (;)-separated list.</li> <li>Optionally enter references for your isolates as a semi-colon (;)-separated list of PubMed ids.</li> <li>You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name.</li> </ul>					
Please paste in tab-delimited text	clude a field header line) Action					
isolate aliases references date received non culture	country region year date_sampled epidemiological year age yr age range SUBMIT					
age_mth_sex disease source MLEE designation server	epidemiology species serogroup genogroup					
penicillin range amox:	illin sulphonamide ceftriaxone mphenicol chloramphenicol range cefotaxime					
cefotaxime_range rifar	icin rifampicin_range ciprofloxacin					
ciprofloxacin_range tetra spectinomycin bioproject a	ycline <u>cefixime azithromycin</u> ession biosample accession ENA run accession					
private_project comments	abcZ adk aroE fumC gdh pdhC					
pgm PorA_VR1 PorA MAST porB NG-MAST tbpB	R2 FetA VR gyrA penA rpoB NG-					
UK325 UK	2020					
meningitis and septicaemia 2 3 4 3	blood <u>Neisseria meningitidis</u> C 8 4 6 5 2 F1-5					
2 3 4 3 UK326 UK	2020					
septicaemia blood	Neisseria meningitidis W					
2 3 4 3	18 4 6 5-1 2 F1-5					
4	•					

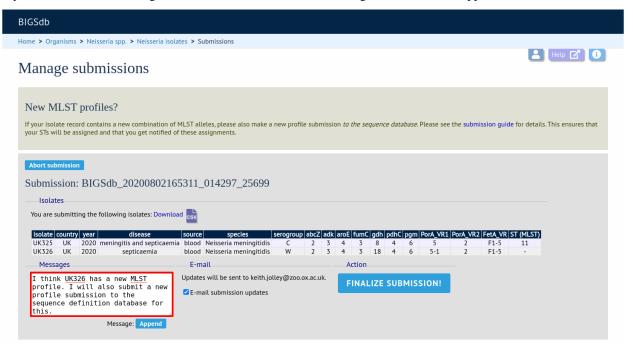
Some basic checks will be performed. These include checking all field values conform to allowed lists or data types. The submission cannot proceed if any checks fail.

BIGSdb	
Home > Organisms > Neisseria spp. > Neisseria isolates > Submissions	
Manage submissions	Help 🕜 🚯
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 (j)-separated list. • Optionally enter references for your isolates as a semi-colon (j)-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)	
isolate aliases references country region year date_sampled ate_received non_culture epidemiological_year age_yr age_range	

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'.



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
020-08-02 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

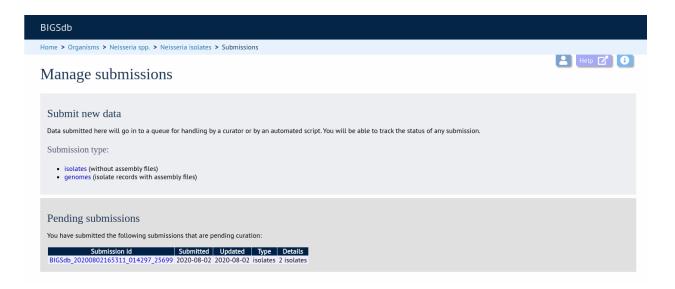
### 16.4.4 Finalize submission

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!'.

E-mail	Action			
Updates will be sent to keith.jol	FINALIZE SUBMISSION!			
E-mail submission updates	E-mail submission updates			

Your submission will then be listed under 'Pending submissions' on your submission page.



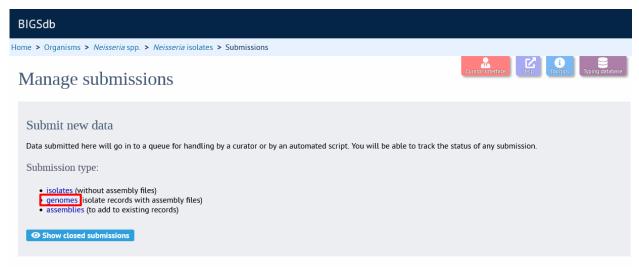
# 16.5 Genome submission

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. You will be able to finalize the submission only after all the assembly files have been uploaded.

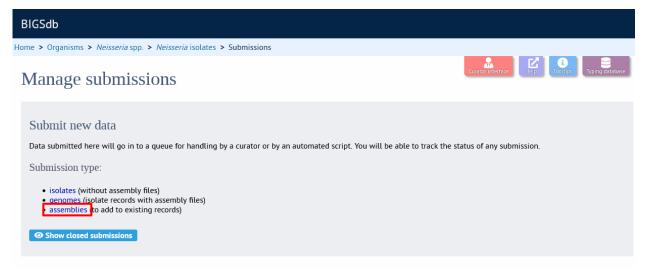
**Note:** When including the filename for your FASTA file containing the genome assembly, please note that Windows will, by default, hide the file extension, e.g. .fas or .fasta. Even if it is hidden in the Windows interface, the file extension is part of the filename and must be included so that the 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 Assembly submission

Genome assemblies can be submitted to add to existing isolate records. These are often old records that have been submitted with just MLST results but whole genome sequencing has been performed later.

### 16.6.1 Start

Click the 'assemblies' link under submission type on the submission management page.



### 16.6.2 Link assembly files to isolate records

You need to tell the system which isolate record should be linked to each assembly that is being uploaded. In order to do this you should prepare a spreadsheet consisting of four columns that you then copy and paste into the web form. The columns are:

- · database id number
- isolate name
- · sequence method
- assembly filename

Both the database id and isolate name are used so that they can be cross-checked to ensure that the correct isolate record has been selected.

^
1
~

Click 'Submit' and the system will check that the id and isolate names correspond tp isolate records that do not currently have assemblies. Provided these match, you will then be prompted to drag-and-drop your genome assemblies on to the web form. A check will also be performed to see if you are the original submitter of the isolate. If you are not, you can still make the submission but should add a message to the curator to confirm why you are adding assemblies for these records.

BIGSdb				^
Home > Organisms > Neisseria spp.	> Neisseria isolates > Submissions			
Manage submissi	ons		Curator Interface	base
	sender for the following isolate ids: 3, 5, 6. t please add a message to the curator to confirm w	hy you are adding assemblies for these i	isolates.	
Supporting files Please upload any supporting file	20221215084433_1675854_5109 es required for curation. Ensure that these are nam u can upload up to 64 MB at a time. If your total su	ed unambiguously or add an explanatory		
Messages	Assemblies			ļ
Message: Append	You are submitting the following isolates: id isolate method filename cont 3 M00242905 Illumina M00242905.fasta 5 M00240227 Illumina M00282207.fasta Missing contig assembly files are shown in red.	igs total length (bp) N50 No sequence No sequence No sequence		
E-mail Updates will be sent to keith.joll ☑ E-mail submission updates	ey@biology.ox.ac.uk.			

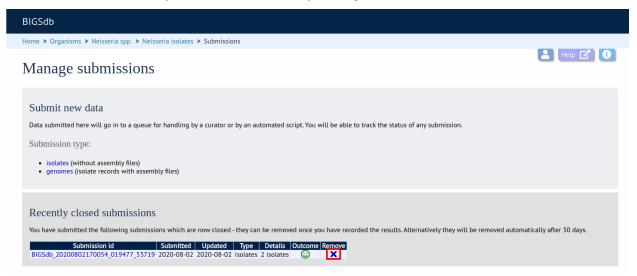
Files are uploaded as you drag-and-drop them. Basic checks wil be performed for sequence length, number of contigs, and N50 values. If values are outside the preferred range you will see a warning for a specific value shown with a pink background. If values are outside the allowed range than the validation will fail and you will need to abort the submission. In the image below, the N50 values have a warning but have not failed the validation.

Uploaded files					
Filename Size Delete	l				
M00282207.fasta 2 MB					
M00242905.fasta 2 MB					
Delete selected files					
Messages	Assemblies				
	You are submitting the follo	-			
	id isolate method 3 M00242905 Illumina M		247	total length (bp	<u></u>
	5 M00242905 Illumina N 5 M00240227 Illumina N			2,133,194	16,320
	6 M00282207 Illumina N		253	2,135,554	17,090
Message: Append	6 MIDU282207 Illumina M	100282207.fasta	232	2,095,969	19,101
	E-mail			Action	
	Updates will be sent to keit	h.jolley@biology.	ox.ac.uk.	FINALIZ	ZE SUBMISSION
	E-mail submission updat	es			
	-				

Once the files have been uploaded and passed validation, add any message to the curator if necessary, e.g. if the original isolate submissions were made by someone else. Click 'Finalize submission'.

# 16.7 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

SEVENTEEN

# **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.

```
curl -i -H "X-PER-PAGE:10" -H "X-OFFSET:0" https://rest.pubmlst.org/db/pubmlst_neisseria_

isolates/isolates
```

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* List groups defined for a classification scheme
- *GET /db/{database}/classification\_schemes/{classification\_scheme\_id}/groups/{group\_id}* List isolates or profiles 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

#### **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 [string] (sequence definition databases only) URI to list of groups
  - 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 - List groups defined for a classification scheme

Sequence definition databases only.

#### **Required route parameters:**

- database [string] Database configuration name
- · classification\_scheme\_id [integer] Classification scheme 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\_seqdef/classification\_schemes/1/groups

Response: Object containing of:

- records [integer] Number of groups
- groups [array] List of URIs to classification group 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)

### 17.3.6 GET /db/{database}/classification\_schemes/{classification\_scheme\_id}/groups/{group\_id - List isolates or profiles belonging to a classification scheme group

#### **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 some of:

- records [integer] Number of isolates or profiles
- isolates (isolate database only) [array] List of *URIs to isolate records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- profiles (sequence definition databases only) [array] List of *URIs to 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)

# 17.3.7 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 (but not on) specified date (ISO 8601 format). Only recognized in sequence definition databases.
- alleles\_updated\_after [date] Include only loci with alleles last modified after (but not on) specified date (ISO 8601 format). Only recognized in sequence definition databases.
- alleles\_added\_reldate [integer] Include only loci with alleles added within the number of days specified. Only recognized in sequence definition databases.
- alleles\_updated\_reldate [integer] Include only loci with alleles last modified within the number of days specified. 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.

## 17.3.8 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)

# 17.3.9 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 (but not on) specified date (ISO 8601 format).
- added\_reldate [integer] Include only alleles added within the specified number of days.
- added\_on [date] Include only alleles added on specified date (ISO 8601 format).
- updated\_after [date] Include only alleles last modified after (but not on) specified date (ISO 8601 format).
- updated\_reldate [integer] Include only alleles updated within the specified number of days.
- 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)

# 17.3.10 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 (but not on) specified date (ISO 8601 format).
- added\_reldate [integer] Include only alleles added within the specified number of days.
- added\_on date] Include only alleles added on specified date (ISO 8601 format).
- updated\_after [date] Include only alleles last modified after (but not on) specified date (ISO 8601 format).
- updated\_reldate [integer] Include only alleles last modified within the specified number of days.
- 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

## 17.3.11 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)

## 17.3.12 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

# 17.3.13 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.

## 17.3.14 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 (but not on) specified date (ISO 8601 format).
- added\_reldate [integer] Count only alleles added within the specified number of days.
- added\_on [date] Count only alleles added on specified date (ISO 8601 format).
- updated\_after [date] Count only alleles last modified after (but not on) specified date (ISO 8601 format).
- updated\_reldate [integer] Count only alleles last modified within the specified number of days.
- 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).

### 17.3.15 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]

### 17.3.16 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 (but not on) specified date (ISO 8601 format).
- added\_reldate [integer] Count only profiles added within the specified number of days.
- added\_on [date] Count only profiles added on specified date (ISO 8601 format).
- updated\_after [date] Count only profiles last modified after (but not on) specified date (ISO 8601 format).
- updated\_reldate [integer] Count only profiles last modified within the specified number of days.
- 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).

## 17.3.17 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 (but not on) specified date (ISO 8601 format). Only recognized in sequence definition databases.
- alleles\_added\_reldate [integer] Include only loci with alleles added within the specified number of days. 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.
- alleles\_updated\_reldate [integer] Include only loci with alleles last modified within the specified number of days. 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.

## 17.3.18 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

## 17.3.19 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 (but not on) specified date (ISO 8601 format).
- added\_reldate [integer] Include only profiles added within the specified number of days.
- added\_on [date] Include only profiles added on specified date (ISO 8601 format).
- updated\_after [date] Include only profiles last modified after (but not on) specified date (ISO 8601 format).
- updated\_reldate [integer] Include only profiles last modified within the specified number of days.
- 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}/schemes/{scheme\_id}/profiles\_csv*.

# 17.3.20 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 (but not on) specified date (ISO 8601 format).
- added\_reldate [integer] Include only profiles added within the specified number of days.
- added\_on [date] Include only profiles added on specified date (ISO 8601 format).
- updated\_after [date] Include only profiles last modified after (but not on) specified date (ISO 8601 format).
- updated\_reldate [integer] Include only profiles last modified within the specified number of days.
- 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

# 17.3.21 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)

# 17.3.22 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:

**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.

# 17.3.23 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"}]}}'
```

# 17.3.24 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 (but not on) specified date (ISO 8601 format).
- added\_reldate [integer] Include only isolates added within the specified number of days.
- 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 (but not on) specified date (ISO 8601 format).
- updated\_reldate [integer] Include only isolates last modified within the specified number of days.
- 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)

### 17.3.25 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

## 17.3.26 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)

## 17.3.27 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)

## 17.3.28 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)

## 17.3.29 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.

## 17.3.30 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.

## 17.3.31 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)

## 17.3.32 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

# 17.3.33 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]

# 17.3.34 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 (but not on) specified date (ISO 8601 format).
- added\_reldate [integer] Include only isolates added within the specified number of days.
- 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 (but not on) specified date (ISO 8601 format).
- updated\_reldate [integer] Include only isolates last modified within the specified number of days.
- 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)

### 17.3.35 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)

## 17.3.36 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

## 17.3.37 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

### 17.3.38 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)

#### 17.3.39 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)

## 17.3.40 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.

## 17.3.41 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

## 17.3.42 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.

# 17.3.43 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)

## 17.3.44 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)

### 17.3.45 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'. Following 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.'

# 17.3.46 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

ExamplerequestURI: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.

# 17.3.47 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

ExamplerequestURI:https://rest.pubmlst.org/db/pubmlst\_neisseria\_seqdef/submissions/BIGSdb\_20151013081836\_14559\_14740

Response: message [string] - 'Submission deleted.'

# 17.3.48 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]

## 17.3.49 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.'

# 17.3.50 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

## 17.3.51 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.'

## 17.3.52 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

## 17.3.53 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')

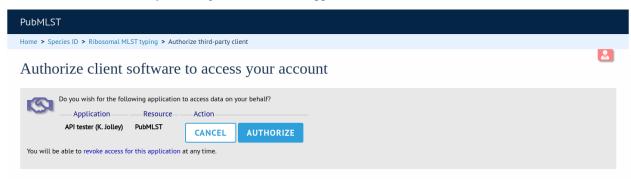
If the application has been registered and has been granted permission to access the specific resource, a JSON response will be returned containing the following parameters:

- 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.

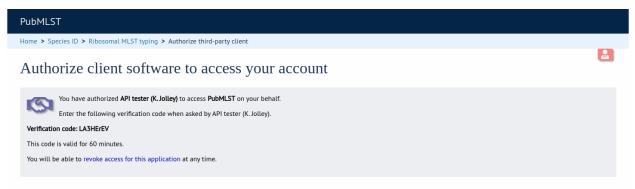
## 17.4.3 Getting user authorization

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:



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. SDrC74ZVI5SYSqY8IWZqrRxnyDnNGVFO
- 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

# EIGHTEEN

# 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

# NINETEEN

# **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

# TWENTY

# DATABASE SCHEMA

- Sequence definition database
- Isolate database

# INDEX

# Α

access control lists, 60 restricting, 61 adding classification groups, 96 isolates. 148 GET /, 421 locus, 69, 71, 72, 77, 101 GET /db, 421 MLST scheme, 94 schemes, 80allele definition 422 records. 225 allele designations 422count, 262 query, 261 423 status, 264 allele sequence 423 identify, 231 allele sequences 442 adding, 126 alleles, 3 list query, 241 retiring, 135 un-retiring, 136 allelic profiles isolates, 273 436 annotation status query, 265 437 API authentication access token, 453 438 accessing protected resources, 454 consumer key, 451 438 request token, 452 session token, 454 439 user authorization, 452 **API** resources 440add submission correspondence, 450 create new submission, 447 451

delete submission record, 449 delete submission supporting file, 451 download alleles in FASTA format, 427 download allelic profiles in CSV (tab-delimited) format, 432 download contigs in FASTA format, 440 download submission supporting file, 450 GET /db/{database}, 422 GET /db/{database}/classification\_schemes, GET /db/{database}/classification\_schemes/{classificat GET /db/{database}/classification\_schemes/{classificat GET /db/{database}/classification\_schemes/{classificat GET /db/{database}/contigs/{contig\_id}, GET /db/{database}/curators, 445 GET /db/{database}/fields, 443 GET /db/{database}/fields/{field},444 GET /db/{database}/genomes, 441 GET /db/{database}/isolates, 435 GET /db/{database}/isolates/{isolate\_id}, GET /db/{database}/isolates/{isolate\_id}/allele\_design GET /db/{database}/isolates/{isolate\_id}/allele\_desigr GET /db/{database}/isolates/{isolate\_id}/allele\_ids, GET /db/{database}/isolates/{isolate\_id}/contigs, GET /db/{database}/isolates/{isolate\_id}/contigs\_fasta GET /db/{database}/isolates/{isolate\_id}/history, DELETE /db/{database}/submissions/{submission\_id}/ffles/{filename}, 451 GET /db/{database}/isolates/{isolate\_id}/schemes/{schemes

439 POST /db/{database}/submissions, 447 GET /db/{database}/loci, 424 POST /db/{database}/submissions/{submission\_id}/files, GET /db/{database}/loci/{locus}, 425 450 POST /db/{database}/submissions/{submission\_id}/messag GET /db/{database}/loci/{locus}/alleles, 426 450GET /db/{database}/loci/{locus}/alleles/{allelaqueick},allele sequence, 427 query allele sequence without 427 GET /db/{database}/loci/{locus}/alleles\_fasta, specifying locus, 428 427 query scheme designations, 434 query scheme sequences, 433 GET /db/{database}/projects, 445 GET /db/{database}/projects/{project\_id}, retrieve allele identifiers, 438 retrieve classification scheme 445 GET /db/{database}/projects/{project\_id}/isolates, information and groups, 422, 423 retrieve contig record, 442 445 GET /db/{database}/schemes, 430 retrieve full allele designation record, GET /db/{database}/schemes/{scheme\_id}, 438 retrieve full allele information, 427 430 GET /db/{database}/schemes/{scheme\_id}/fields/feftielet/feftielet 431 431 GET /db/{database}/schemes/{scheme\_id}/loci, retrieve isolate record, 436 431 retrieve isolate update history, 440 GET /db/{database}/schemes/{scheme\_id}/profilesetrieve list of allele designations, 437 retrieve list of alleles defined for a 431 GET /db/{database}/schemes/{scheme\_id}/profiles/{phoofid.e42id}, retrieve list of contigs, 439 433 GET /db/{database}/scheme\_id}/profilesetsive list of curators, 445 432 retrieve list of groups for a GET /db/{database}/sequences, 429 classification scheme, 423 GET /db/{database}/submissions, 446 retrieve list of isolate provenance GET /db/{database}/submissions/{submission\_id}, field descriptions, 443 448 retrieve list of isolate records, 435 GET /db/{database}/submissions/{submission\_id}/ftilitesve list of isolate records that 450 have genome assemblies, 441 GET /db/{database}/submissions/{submission\_id}r/eftilies/effilename}isolates belonging to a project, 445 450 GET /db/{database}/submissions/{submission\_id}/memory.445 449 retrieve list of scheme allele GET /db/{database}/users/{user\_id},444 identifiers, 439 retrieve list of submissions, 446 get summary of defined sequences, 429 list allelic profiles defined for retrieve list of supporting files scheme, 431 uploaded for submission, 450 list classification schemes, 422 retrieve locus record, 425 retrieve project information, 445 list database resources, 422 list loci, 424 retrieve scheme allele designation list schemes, 430 records, 439 list site resources, 421 retrieve scheme information, 430 POST /db/{database}/isolates/search, 442 retrieve scheme loci, 431 retrieve specific allelic profile POST /db/{database}/loci/{locus}/sequence, 427 record, 433 POST /db/{database}/schemes/{scheme\_id}/designationesve submission correspondence, 449 retrieve submission record, 448 434 POST /db/{database}/schemes/{scheme\_id}/sequencetrieve user information, 444 433 retrieve values set for a provenance POST /db/{database}/sequence, 428 field, 444

search isolate database, 442
upload submission supporting file, 450
assembly stats, 209
auto allele definer, 207
stop, 211
automated assignment
scheme profiles, 95
autotagger, 203
stop, 211

# В

batch profile definitions query, 251 batch uploading contigs - multiple isolates, 166 **BLAST, 307 BLAST** caches refreshing, 122 bookmarks. 276 breakdown provenance field, 320 sequence bin, 374 two-field, 377 browse isolates, 257 scheme profiles, 242 **BURST**, 313

# С

caching schemes, 64 classification groups, 4 adding, 96 client authorization RESTful interface, 121 client databases, 89 clustering core genome, 99 codon usage, 317 Locus Explorer, 359 composite fields, 111 config-specific downloads, 123 configuration export, 119 configuration settings validation, 118 core genome clustering, 99 count allele designations, 262 sequence tags, 267

# D

```
defining
exemplar alleles, 206
deleting
isolates, 153, 156
profiles, 144
downloads
config-specific, 123
```

# E

```
exemplar alleles
defining, 206
export
configuration, 119
extended attributes
locus, 78
provenance fields, 114
```

# F

filters, 272

# G

```
gene presence;, 330
Genome Comparator, 336
genome filtering, 106
    in silico hybridization, 108
    in silico PCR, 106
geographic coordinates, 181
geographic point lookup values
    populating, 212
GrapeTree
    minimum-spanning trees, 344
groups
    scheme, 86
```

# Η

hosts mapping, 63

# I

identify
 allele sequence, 231
 sequence type, 233
in silico hybridization, 108
in silico PCR, 106
in silico PCR analysis, 348
isolate
 records, 221
isolate aliases, 160
isolate record
 options, 286
isolates
 adding, 148

```
allelic profiles, 273
browse, 257
deleting, 153, 156
query, 258
retiring, 157
un-retiring, 158
updating, 152, 155
uploading contigs, 163
iTOL
phylogenetic trees, 351
```

# K

kiosk mode, 44

# L

LINcode schemes, 102 linking remote contigs, 169 list query alleles, 241 loci.3 modifying display, 290 locus adding, 69, 71, 72, 77, 101 adding; copying existing record, 76 extended attributes, 78 Locus Explorer, 356 codon usage, 359 polymorphic sites, 356 translated sequences, 360 locus positions setting, 110

# Μ

main results table options. 284 mapping hosts, 63 Microreact spatio-temporaral trees, 361 minimum-spanning trees GrapeTree, 344 PhyloViz online, 364 MLST, 3 MLST scheme adding, 94 mod\_perl, 64 modifying display loci. 290 schemes, 290

# 0

offline curation auto allele definer, 207 autotagger, 203 options, 282 isolate record, 286 main results table, 284 provenance fields, 288 query, 289

# Ρ

partitioning sets, 65 passwords setting, 61 setting; first user, 61 performance caching schemes, 64 mod\_perl, 64 permissions, 57 locus curation, 59 scheme curation, 59 phylogenetic trees iTOL, 351 PhyloViz online minimum-spanning trees, 364 plugins enabling, 62 polymorphic sites Locus Explorer, 356 polymorphisms, 367 populating geographic point lookup values, 212 profile records, 227 profiles, 4 deleting, 144 retiring, 145 un-retiring, 146 updating, 143 projects, 174 provenance field breakdown. 320 provenance fields extended attributes, 114 options, 288 publications, 161, 279

# Q

```
query
   allele designations, 261
   batch profile definitions, 251
   isolates, 258
   options, 289
   scheme profiles, 244
   sequence tags, 269
   ST definitions from allelic profiles, 248
```

# R

```
records
    allele definition, 225
    isolate, 221
    profile, 227
    sequence bin, 228
    sequence tag, 226
refreshing
    BLAST caches, 122
RESTful interface
    client authorization, 121
retiring
    alleles, 135
    isolates, 157
    profiles, 145
rMLST
    species identification, 370
    storing analysis results, 210
```

# S

scheme groups, 86 scheme profiles automated assignment, 95 browse, 242 query, 244 schemes, 3 adding, 80 caching, 64 modifying display, 290 sequence bin breakdown, 374 query, 266 records, 228 sequence similarity determining, 254 sequence tag records, 226 sequence tags, 4 count, 267query, 269 sequence type identify, 233 sets, 4 partitioning, 65 sparsely-populated fields, 42 spatio-temporaral trees Microreact, 361 species identification rMLST, 370 ST definitions from allelic profiles query, 248 stop auto allele definer, 211

autotagger, 211 storing analysis results rMLST, 210

# Т

translated sequences Locus Explorer, 360 two-field breakdown, 377

# U

```
un-retiring
    alleles, 136
    isolates, 158
    profiles, 146
unique combinations, 380
updates
    disabling, 62
updating
    isolates, 152, 155
    profiles, 143
uploading contigs
    isolates, 163
user groups, 57
user projects, 294
user types, 57
users
    adding, 125
```