
BIGSdb Documentation

Release 1.11.0

Keith Jolley

July 02, 2015

| | | |
|----------|---|-----------|
| 1 | Concepts and terms | 3 |
| 1.1 | BIGSdb | 3 |
| 1.2 | Loci | 3 |
| 1.3 | Alleles | 3 |
| 1.4 | Schemes | 4 |
| 1.5 | Profiles | 4 |
| 1.6 | Sequence tags | 4 |
| 1.7 | Sets | 4 |
| 2 | BIGSdb dependencies | 7 |
| 2.1 | Required packages | 7 |
| 3 | Installation and configuration of BIGSdb | 9 |
| 3.1 | Software installation | 9 |
| 3.2 | Configuring PostgreSQL | 10 |
| 3.3 | Site-specific configuration | 11 |
| 3.4 | Setting up the offline job manager | 11 |
| 3.5 | Setting up the submission system | 12 |
| 3.6 | Periodically delete temporary files | 13 |
| 3.7 | Prevent preference database getting too large | 13 |
| 3.8 | Purging old jobs from the jobs database | 13 |
| 3.9 | Log file rotation | 14 |
| 3.10 | Upgrading BIGSdb | 14 |
| 3.11 | Running the BIGSdb RESTful interface | 14 |
| 4 | Database setup | 17 |
| 4.1 | Creating databases | 17 |
| 4.2 | Database-specific configuration | 18 |
| 4.3 | XML configuration attributes used in config.xml | 18 |
| 4.4 | User authentication | 25 |
| 4.5 | Setting up the admin user | 26 |
| 4.6 | Updating PubMed citations | 26 |
| 5 | Administrator's guide | 27 |
| 5.1 | Types of user | 27 |
| 5.2 | User groups | 27 |
| 5.3 | Curator permissions | 27 |
| 5.4 | Locus and scheme permissions (sequence definition database) | 30 |

| | | |
|----------|---|------------|
| 5.5 | Controlling access | 32 |
| 5.6 | Setting user passwords | 32 |
| 5.7 | Setting the first user password | 33 |
| 5.8 | Enabling plugins | 34 |
| 5.9 | Temporarily disabling database updates | 34 |
| 5.10 | Host mapping | 35 |
| 5.11 | Improving performance | 35 |
| 5.12 | Dataset partitioning | 37 |
| 5.13 | Adding new loci | 40 |
| 5.14 | Defining locus extended attributes | 52 |
| 5.15 | Defining schemes | 55 |
| 5.16 | Organizing schemes into hierarchical groups | 62 |
| 5.17 | Setting up client databases | 65 |
| 5.18 | Rule-based sequence queries | 71 |
| 5.19 | Workflow for setting up a MLST scheme | 73 |
| 5.20 | Defining new loci based on annotated reference genome | 74 |
| 5.21 | Genome filtering | 76 |
| 5.22 | Setting locus genome positions | 80 |
| 5.23 | Defining composite fields | 85 |
| 5.24 | Extended provenance attributes (lookup tables) | 88 |
| 5.25 | Sequence bin attributes | 91 |
| 5.26 | Checking external database configuration settings | 93 |
| 5.27 | Exporting table configurations | 94 |
| 5.28 | Authorizing third-party client software to access authenticated resources | 96 |
| 6 | Curator's guide | 99 |
| 6.1 | Adding new sender details | 99 |
| 6.2 | Adding new allele sequence definitions | 100 |
| 6.3 | Updating and deleting allele sequence definitions | 108 |
| 6.4 | Updating locus descriptions | 111 |
| 6.5 | Adding new scheme profile definitions | 115 |
| 6.6 | Updating and deleting scheme profile definitions | 119 |
| 6.7 | Adding isolate records | 121 |
| 6.8 | Updating and deleting single isolate records | 125 |
| 6.9 | Batch updating multiple isolate records | 130 |
| 6.10 | Deleting multiple isolate records | 132 |
| 6.11 | Uploading sequence contigs linked to isolate records | 133 |
| 6.12 | Automated web-based sequence tagging | 138 |
| 6.13 | Projects | 141 |
| 6.14 | Isolate record versioning | 146 |
| 7 | Curating submitted data | 149 |
| 7.1 | Alleles | 150 |
| 7.2 | Profiles | 157 |
| 7.3 | Isolates | 163 |
| 8 | Offline curation tools | 169 |
| 8.1 | Automated offline sequence tagging | 169 |
| 8.2 | Automated offline allele definition | 171 |
| 8.3 | Cleanly interrupting offline curation | 173 |
| 8.4 | Uploading contigs from the command line | 173 |
| 9 | Definition downloads | 175 |
| 9.1 | Allele sequence definitions | 175 |
| 9.2 | Scheme profile definitions | 179 |

| | |
|---|------------|
| 10 Data records | 181 |
| 10.1 Isolate records | 181 |
| 10.2 Allele definition records | 185 |
| 10.3 Sequence tag records | 186 |
| 10.4 Profile records | 187 |
| 10.5 Sequence bin records | 188 |
| 11 Querying data | 189 |
| 11.1 Querying sequences to determine allele identity | 189 |
| 11.2 Searching for specific allele definitions | 192 |
| 11.3 Browsing scheme profile definitions | 197 |
| 11.4 Querying scheme profile definitions | 199 |
| 11.5 Investigating allele differences | 200 |
| 11.6 Querying isolate data | 204 |
| 11.7 Querying by allelic profile | 212 |
| 11.8 Retrieving list of isolates or profiles | 215 |
| 11.9 Retrieving isolates by linked publication | 216 |
| 11.10 User-configurable options | 219 |
| 12 Data analysis plugins | 231 |
| 12.1 Locus explorer | 231 |
| 12.2 Field breakdown | 236 |
| 12.3 Two field breakdown | 239 |
| 12.4 Scheme and allele breakdown | 242 |
| 12.5 Sequence bin breakdown | 246 |
| 12.6 Genome comparator | 249 |
| 12.7 BLAST | 257 |
| 12.8 BURST | 261 |
| 12.9 Codon usage | 265 |
| 12.10 Unique combinations | 268 |
| 12.11 Polymorphisms | 270 |
| 12.12 Presence/absence | 272 |
| 12.13 Tag status | 275 |
| 13 Data export plugins | 279 |
| 13.1 Isolate record export | 279 |
| 13.2 Sequence export | 282 |
| 13.3 Contig export | 284 |
| 14 Submitting data using the submission system | 289 |
| 14.1 Registering a user account | 289 |
| 14.2 Allele submission | 290 |
| 14.3 Profile submission | 299 |
| 14.4 Isolate submission | 306 |
| 15 RESTful Application Programming Interface (API) | 315 |
| 15.1 Passing optional parameters | 315 |
| 15.2 Resources | 315 |
| 15.3 Authentication | 329 |
| 16 Frequently asked questions (FAQs) | 335 |
| 16.1 General | 335 |
| 16.2 Installation | 335 |
| 16.3 Administration | 336 |

| | |
|--------------------------------------|------------|
| 17 Appendix | 337 |
| 17.1 Query operators | 337 |
| 17.2 Sequence tag flags | 337 |
| 17.3 Allele sequence flags | 338 |
| 18 Database schema | 341 |

Gene-by-gene population annotation and analysis

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

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

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

Concepts and terms

1.1 BIGSdb

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

1.2 Loci

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

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

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

1.3 Alleles

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

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

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

1.4 Schemes

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

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

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

1.5 Profiles

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

1.6 Sequence tags

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

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

1.7 Sets

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

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

BIGSdb dependencies

2.1 Required packages

BIGSdb requires a number of software components to be installed:

2.1.1 Linux packages

- Apache2 web server with mod_perl2
- PostgreSQL database
- Perl 5.10+
- BioPerl
- BLAST+
- EMBOSS
 - infoalign - use to extract alignment stats in Genome Comparator.
 - sixpack - used to translate sequences in multiple reading frames.
 - stretcher - used for sequence alignment in allele query.
- Ipcress - part of exonerate package - used to simulate PCR reactions which can be used to filter the genome to predicted amplification products.
- Xvfb - X virtual framebuffer - needed to support SplitsTree in command line mode as used in Genome Comparator.

2.1.2 Perl modules

These are included with most Linux distributions.

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

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

2.1.3 Optional packages

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

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

Installation and configuration of BIGSdb

3.1 Software installation

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

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

- [Software requirements](#)
- Download from [SourceForge.net](#) or [GitHub](#).

1. Unpack the distribution package in a temporary directory:

```
gunzip bigsdB_1.x.x.tar.gz
tar xvf bigsdB_1.x.x.tar
```

2. Copy the `bigsdB.pl` and `bigscurate.pl` scripts to a subdirectory of your web server's cgi-bin directory. Make sure these are readable and executable by the web server daemon.
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 contents of the javascript directory to a javascript directory within the web root tree, i.e. accessible from http://your_website/javascript/.
5. Copy the contents of the css directory to a css directory within the web root tree, i.e. accessible from http://your_website/css/.
6. Copy the contents of the fonts directory to a fonts directory within the web root tree, i.e. accessible from http://your_website/fonts/.
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 privileges. First you will need to log in as the postgres user:

```
sudo su postgres
```

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

```
createuser apache
```

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

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

10. Create PostgreSQL databases called bigsdb_auth, bigsdb_prefs and bigsdb_refs using the scripts in the sql directory. Create the database using the createdb command and set up the tables using the psql command.

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

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

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

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

3.2 Configuring PostgreSQL

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

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

You will need to edit the pg_hba.conf and pg_ident.conf files. These are found somewhere like /etc/postgresql/9.1/main/

pg_hba.conf

```
# Database administrative login by UNIX sockets
local all postgres ident map=mymap

# TYPE DATABASE USER CIDR-ADDRESS METHOD

# "local" is for Unix domain socket connections only
local all all ident map=mymap
# IPv4 local connections:
host all all 127.0.0.1/32 md5
# IPv6 local connections:
host all all ::1/128 md5
```

pg_ident.conf

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

| | | |
|-------|--------|--------|
| mymap | bigbdb | bigbdb |
| mymap | bigbdb | apache |

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

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

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

See [Tuning Your PostgreSQL Server](#) for more details.

Restart PostgreSQL after any changes, e.g.

```
/etc/init.d/postgresql restart
```

3.3 Site-specific configuration

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

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

3.4 Setting up the offline job manager

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

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

```
sudo useradd -s /bin/sh bigbdb
```

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

```
createuser bigbdb [no need for special privileges]
createdb bigbdb_jobs
psql -f jobs.sql bigbdb_jobs
```

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

```
psql
ALTER ROLE bigbdb WITH PASSWORD 'bigbdb';
```

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

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

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

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

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

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

Add the following to `/etc/crontab`:

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

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

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

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

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

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

3.5 Setting up the submission system

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

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). 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 ln -s /var/submissions /var/www
```

2. Set the `submission_dir` location in `bigsdb.conf`.
3. Set the `smtp_server` in `bigsdb.conf` to the IP or DNS name of your organisation's SMTP relay. Depending on how your E-mail system is configured, you may be able to use the localhost address (127.0.0.1).
4. Make sure the `curate_script` and `query_script` values are set in `bigsdb.conf`. These point to the web-accessible location of the web scripts and are required to allow curators to be directed between the web interfaces as needed.

5. Set `submissions="yes"` in the system tag of the *database config.xml* file of each database for which submissions should be enabled.

3.6 Periodically delete temporary files

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

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

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

#Remove .jnlp files from web tree older than 1 day
find /var/www/tmp/ -name '*.jnlp' -type f -mmin +1440 -exec rm -f {} \; 2>/dev/null

#Remove other tmp files from web tree older than 1 week
find /var/www/tmp/ -type f -mmin +10080 -exec rm -f {} \; 2>/dev/null
```

3.7 Prevent preference database getting too large

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

```
#Prevent prefs database getting too large
00 18 * * 0 postgres psql -c "DELETE FROM guid WHERE last_accessed < NOW() - INTERVAL '6 months'"
```

3.8 Purging old jobs from the jobs database

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

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

3.9 Log file rotation

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

```
/var/log/bigsdb.log {
    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
}
```

3.10 Upgrading BIGSdb

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

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

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

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

3.11 Running the BIGSdb RESTful interface

BIGSdb has an Application Programming Interface (API) that allows third-party applications to access the data within the databases. The script that runs this is called bigsrest.pl. This is a Dancer2 application that can be run using a wide range of options, e.g. as a stand-alone script, using Perl webserver 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 web-server 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.

3.11.1 Proxying the API to use a standard web port

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

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

  ProxyPass / http://rest.pubmlst.org:5000/
  ProxyPassReverse / http://rest.pubmlst.org:5000/

  <Proxy *>
    Order allow,deny
    Allow from all
  </Proxy>

  ErrorLog /var/log/apache2/rest.pubmlst.org-error.log
  CustomLog /var/log/apache2/rest.pubmlst.org-access.log common
</VirtualHost>
```

Database setup

There are two types of BIGSdb database:

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

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

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

4.1 Creating databases

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

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

```
sudo su postgres
```

then

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

Create an isolate database the same way:

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

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

```
psql bigsdb_test_isolates
```

and alter the isolate table::

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

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

4.2 Database-specific configuration

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

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

4.3 XML configuration attributes used in config.xml

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

4.3.1 Isolate database XML attributes

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

```
<db>
```

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

```
<system>
```

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

- **db**
 - Name of database on system.
- **dbtype**
 - Type of database: either ‘isolates’ or ‘sequences’.
- **description**
 - Description of database used throughout interface.
- **align_limit**
 - Overrides the sequence export record alignment limit in the Sequence Export plugin. Default: ‘200’.
- **all_plugins**
 - Enable all appropriate plugins for database: either ‘yes’ or ‘no’, default ‘no’.
- **annotation**
 - Semi-colon separated list of accession numbers with descriptions (separated by a |), eg. ‘AL157959|Z2491;AM421808|FAM18;NC_002946|FA 1090;NC_011035|NCCP11945;NC_014752|020-06’. Currently used only by Genome Comparator plugin.
- **codon_usage_limit**
 - Overrides the record limit for the Codon Usage plugin. Default: ‘500’.
- **contig_analysis_limit**
 - Overrides the isolate number limit for the Contig Export plugin. Default: ‘1000’.
- **curate_only**
 - Set to ‘yes’ to hide field on an isolate information page in the standard interface. The field will be visible if the page is accessed via the curator’s interface (version 1.10.0+).
- **curate_path_includes**
 - Partial path of the bigscurate.pl script used to curate the database. See user authentication.
- **curate_script**
 - Relative web path to curation script. Default ‘bigscurate.pl’ (version 1.11+).
 - This is only needed if automated submissions are enabled. If bigscurate.pl is in a different directory from bigsdb.pl, you need to include the whole web path, e.g. /cgi-bin/private/bigsdb/bigscurate.pl.
- **default**
 - Default value. This will be entered automatically in the web form but can be overridden.
- **default_access**
 - The default access to the database configuration, either ‘allow’ or ‘deny’. If ‘allow’, then specific users can be denied access by creating a file called ‘users.deny’ containing usernames (one per line) in the configuration directory. If ‘deny’ then specific users can be allowed by creating a file called ‘users.allow’ containing usernames (one per line) in the configuration directory. See [default access](#).
- **default_seqdef_config**
 - Isolate databases only: Name of the default seqdef database configuration used with this database. Used to automatically fill in details when adding new loci.
- **default_seqdef_dbase**

- Isolate databases only: Name of the default seqdef database used with this database. Used to automatically fill in details when adding new loci.
- default_seqdef_script
 - Isolate databases only: URL of BIGSdb script running the seqdef database (default: `/cgi-bin/bigsdb/bigsdb.pl`).
- fieldgroup1 - fieldgroup10
 - Allows multiple fields to be queried as a group. Value should be the name of the group followed by a colon (:) followed by a comma-separated list of fields to group, e.g. `identifiers:id,strain,other_name`.
- genome_comparator_limit
 - Overrides the isolate number limit for the Genome Comparator plugin. Default: `'1000'`.
- genome_comparator_max_ref_loci
 - Overrides the limit on number of loci allowed in a reference genome. Default: `'10000'`.
- hide_unused_schemes
 - Sets whether a scheme is shown in a main results table if none of the isolates on that page have any data for the specific scheme: either `'yes'` or `'no'`, default `'no'`.
- host
 - Host name/IP address of machine hosting isolate database, default `'localhost'`.
- job_priority
 - Integer with default job priority for offline jobs (default:5).
- job_quota
 - Integer with number of offline jobs that can be queued or currently running for this database.
- labelfield
 - Field that is used to describe record in isolate info page, default `'isolate'`.
- locus_superscript_prefix
 - Superscript the first letter of a locus name if it is immediately following by an underscore, e.g. `f_abcZ` would be displayed as `fabcZ` within the interface: must be either `'yes'` or `'no'`, default `'no'`. This can be used to designate gene fragments (or any other meaning you like).
- maindisplay_aliases
 - Default setting for whether isolates aliases are displayed in main results tables: either `'yes'` or `'no'`, default `'no'`. This setting can be overridden by individual user preferences.
- noshow
 - Comma-separated list of fields not to use in breakdown statistic plugins.
- no_publication_filter
 - Isolate databases only: Switches off display of publication filter in isolate query form by default: either `'yes'` or `'no'`, default `'no'`.
- only_sets
 - Don't allow option to view the `'whole database'` - only list sets that have been defined: either `'yes'` or `'no'`, default `'no'`.
- password

- Password for access to isolates database, default ‘remote’.
- port
 - Port number that the isolate host is listening on, default ‘5432’.
- privacy
 - Displays E-mail address for sender in isolate information page if set to ‘no’. Default ‘yes’.
- query_script
 - Relative web path to bigsdb script. Default ‘bigsdb.pl’ (version 1.11+).
 - This is only needed if automated submissions are enabled. If bigsdb.pl is in a different directory from bigscurate.pl, you need to include the whole web path, e.g. /cgi-bin/bigsdb/bigsdb.pl.
- read_access
 - Describes who can view data: either ‘public’ for everybody or ‘authenticated_users’ for anybody who has been able to log in. Default ‘public’.
- script_path_includes
 - Partial path of the bigsdb.pl script used to access the database. See [user authentication](#).
- seqbin_size_threshold
 - Sets the size values in Mbp to enable for the [seqbin filter](#).
 - Example: seqbin_size_threshold=”0.5,1,2,4”.
- seq_export_limit
 - Overrides the sequence export limit (records x loci) in the Sequence Export plugin. Default: ‘1000000’.
- sets
 - Use [sets](#): either ‘yes’ or ‘no’, default ‘no’.
- set_id
 - Force the use of a specific set when accessing database via this XML configuration: Value is the name of the set.
- start_id
 - Defines the minimum record id to be used when uploading new isolate records. This can be useful when it is anticipated that two databases may be merged and it would be easier to do so if the id numbers in the two databases were different. Default: ‘1’.
- submissions
 - Enable automated submission system: either ‘yes’ or ‘no’, default ‘no’ (version 1.11+).
 - The curate_script and query_script paths should also be set, either in the bigsdb.conf file (for site-wide configuration) or within the system attribute of config.xml.
- tblastx_tagging
 - Sets whether tagging can be performed using TBLASTX: either ‘yes’ or ‘no’, default ‘no’.
- user
 - Username for access to isolates database, default ‘apache’.
- use_temp_scheme_table

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

`<field>`

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

```
<field type="text" required="no" length="40" maindisplay="no"
  web="http://somewebsite.com/cgi-bin/script.pl?id=[?]" optlist="yes">epidemiology
<optlist>
  <option>carrier</option>
  <option>healthy contact</option>
  <option>sporadic case</option>
  <option>endemic</option>
  <option>epidemic</option>
  <option>pandemic</option>
</optlist>
</field>
```

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

- Maximum value for integer types. Special values such as CURRENT_YEAR can be used.
- min
 - Minimum value for integer types.
- optlist
 - Sets if this field has a list of allowed values, default ‘no’. Surround each option with an <option> tag.
- regex
 - Regular expression used to constrain field values, e.g. regex=”^[A-Z].*\$” forces the first letter of the value to be capitalized.
- required
 - Sets if data is required for this field, ‘yes’ or ‘no’, default ‘yes’.
- userfield
 - Select if you want this field to have its own dropdown filter box of users (populated from the users table): ‘yes’ or ‘no’, default ‘no’.
- web
 - URL that will be used to hyperlink field values. If [?] is included in the URL, this will be substituted for the actual field value.

Special values

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

- CURRENT_YEAR: the 4 digit value of the current year

```
<sample>
```

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

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

4.3.2 Sequence definition database XML attributes

Required attributes are in **bold**.

```
<db>
```

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

```
<system>
```

- **authentication**
 - Method of authentication: either ‘builtin’ or ‘apache’. See [user authentication](#).
- **db**
 - Name of database on system.
- **dbtype**
 - Type of database: either ‘isolates’ or ‘sequences’.

- **description**
 - Description of database used throughout interface.
- **align_limit**
 - Overrides the sequence export record alignment limit in the Sequence Export plugin. Default: '200'.
- **allele_comments**
 - Enable comments on allele sequences: either 'yes' or 'no', default 'no'.
 - This is not enabled by default to discourage the practice of adding isolate information to allele definitions (this sort of information belongs in an isolate database).
- **allele_flags**
 - Enable flags to be set for alleles: either 'yes' or 'no', default 'no'.
- **curate_path_includes**
 - Partial path of the bigscure.pl script used to curate the database. See [user authentication](#).
- **curate_script**
 - Relative web path to curation script. Default 'bigscure.pl' (version 1.11+).
 - This is only needed if automated submissions are enabled. If bigscure.pl is in a different directory from bigsdb.pl, you need to include the whole web path, e.g. /cgi-bin/private/bigsdb/bigscure.pl.
- **diploid**
 - Allow IUPAC 2-nucleotide ambiguity codes in allele definitions for use with diploid typing schemes: either 'yes' or 'no', default 'no'.
- **disable_seq_downloads**
 - Prevent users or curators from downloading all alleles for a locus (admins always can). 'yes' or 'no', default 'no'.
- **job_priority**
 - Integer with default job priority for offline jobs (default:5).
- **job_quota**
 - Integer with number of offline jobs that can be queued or currently running for this database.
- **materialized_views**
 - Enable materialized views: either 'yes' or 'no', default 'no'.
- **profile_submissions**
 - Enable profile submissions (automated submission system): either 'yes' or 'no', default 'no' (version 1.11+).
 - To enable, you will also need to set submissions="yes". By default, profile submissions are disabled since generally new profiles should be accompanied by representative isolate data, and the profile can be extracted from that.
- **query_script**
 - Relative web path to bigsdb script. Default 'bigsdb.pl' (version 1.11+).
 - This is only needed if automated submissions are enabled. If bigsdb.pl is in a different directory from bigscure.pl, you need to include the whole web path, e.g. /cgi-bin/bigsdb/bigsdb.pl.
- **read_access**

- Describes who can view data: either ‘public’ for everybody, or ‘authenticated_users’ for anybody who has been able to log in. Default ‘public’.
- script_path_includes
 - Partial path of the bigsdb.pl script used to access the database. See [user authentication](#).
- seq_export_limit
 - Overrides the sequence export limit (records x loci) in the Sequence Export plugin. Default: ‘1000000’.
- sets
 - Use [sets](#): either ‘yes’ or ‘no’, default ‘no’.
- set_id
 - Force the use of a specific set when accessing database via this XML configuration: Value is the name of the set.
- submissions
 - Enable automated submission system: either ‘yes’ or ‘no’, default ‘no’ (version 1.11+).
 - The curate_script and query_script paths should also be set, either in the bigsdb.conf file (for site-wide configuration) or within the system attribute of config.xml.
- webroot
 - URL of web root, which can be relative or absolute. The bigsdb.css stylesheet file should be located in this directory. Default ‘/’.

4.4 User authentication

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

4.4.1 Apache authentication

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

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

- script_path_includes: the BIGSdb script path must contain the value set.
- curate_path_includes: the BIGSdb curation script path must contain the value set.

For public databases, the ‘script_path_includes’ attribute need not be set.

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

4.4.2 Built-in authentication

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

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

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

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

4.5 Setting up the admin user

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

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

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

4.6 Updating PubMed citations

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

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

| | |
|--|--|
| <code>pubmlst_bigsdb_neisseria_isolates</code> | <code>refs</code> |
| <code>pubmlst_bigsdb_neisseria_seqdef</code> | <code>profile_refs, sequence_refs, locus_refs</code> |

The script can be called as follows:

```
getrefs.pl getrefs.conf
```

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

Administrator's guide

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

5.1 Types of user

There are four types of user in BIGSdb:

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

5.2 User groups

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

5.3 Curator permissions

Individual permissions can be set for each curator:

- `disable_access` - if set to true, this user is completely barred from access.
- `modify_users` - allowed to add or modify user records. They can change the status of users, but can not revoke admin 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_schemes` - allowed to add or modify schemes.
- `modify_sequences` - allowed to add sequences to the sequence bin (for isolate databases) or new allele definitions (for sequence definition databases).
- `modify_experiments` - define new experiments that can be used to group contigs uploaded to the sequence bin.
- `modify_isolates` - allowed to add or modify isolate records.
- `modify_projects` - allowed to create projects, modify their descriptions and add or remove isolate records to these.
- `modify_composites` - allowed to add or modify composite fields (fields made up of other fields, including scheme fields defined in external databases). Composite fields involve defining regular expressions that are evaluated by Perl - this can be dangerous so this permission should be granted with discretion.
- `modify_field_attributes` - allow user to create or modify secondary field attributes (lookup tables) for isolate record fields.
- `modify_value_attributes` - allow user to add or modify secondary field values for isolate record fields.
- `modify_probes` - allow user to define PCR or hybridization reactions to filter tag scanning.
- `tag_sequences` - allowed to tag sequences with locus information.
- `designate_alleles` - allowed to manually designate allele numbers for isolate records.
- `modify_profiles` - allowed to add or modify scheme profiles (only used in a sequence definitions database).

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


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* PubMLST

 Add, update or delete records

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

 Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-----------------------------------|-----|---------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attributes | + | ++ | ? | Define additional attributes to associate with values of a particular isolate record field. |

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#)

Set curator permissions

Select curator(s) Action

Bray, James
 Christodoulides, Myron
 Clemence, Marianne
 Diallo, Kanny
 Exley, Rachel
 Feavers, Ian
 Goodyer-Sait, Lily
 Halkilahti, Jani

All None

Select

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

Logged in: Keith Jolley (keith). [Log out](#) | [Change password](#) [Help](#)

Set curator permissions

Select curator(s)

Bray, James
Christodoulides, Myron
Clemence, Marianne
Diallo, Kanny
Exley, Rachel
Feavers, Ian
Goodyer-Sait, Lily
Halkilahti, Jani

All None

Action

Select

Check the boxes for the required permissions. Users with a status of ‘submitter’ have a restricted list of allowed permissions that can be selected.

Update permissions

| Permission | Curator | | All/None |
|-------------------------|-------------------------------------|-------------------------------------|--------------------------|
| | Clemence, Marianne | Diallo, Kanny | |
| modify users | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| modify isolates | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| modify projects | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| modify sequences | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| tag sequences | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| designate alleles | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| modify usergroups | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| set user passwords | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| modify loci | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| modify schemes | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| modify composites | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| modify field attributes | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| modify value attributes | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| modify probes | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| modify experiments | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| delete all | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| disable access | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| All/None | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Action

Update

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’:

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* locus/sequence definitions

 Add, update or delete records

| Record type | Add | Batch Add | Update or delete | Comments |
|---|-----|------------|--------------------------------------|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rplF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

Select the curator from the list:

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Batch update locus curator access records

Select user

The user status must also be [set to curator](#) for permissions to work.

users:

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:

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#), [Log out](#) | [Change password](#)

Batch update locus curator access records

User: **Keith Jolley**
 Select values to enable or disable and then click the appropriate arrow button.

Select loci

| Available | | Selected |
|----------------|---|----------|
| FHbp_segment_D | | 'porA |
| FHbp_segment_E | | NEIS1364 |
| FetA_VR | | PorA_VR1 |
| NEIS0001 | | PorA_VR2 |
| NEIS0004 | > | abcZ |
| NEIS0005 | < | adk |
| NEIS0006 | | aroE |
| NEIS0007 | | aspA |
| NEIS0008 | | carB |
| NEIS0009 | | dhpS |
| NEIS0010 | | fumC |
| NEIS0011 | | gdh |

☒ Hide curator name from public view

[Back to main](#)

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

5.5 Controlling access

5.5.1 Restricting particular configurations to specific user accounts

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

In the config.xml, add the following directive:

```
default_access="deny"
```

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

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


```
default_access="allow"
```

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

5.6 Setting user passwords

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


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

 **Database configuration**

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

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

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

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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)

Set user password

User: Jolley, Keith (keith) ▼

New 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 results',
        category  => 'Cluster',
        buttontext => 'BURST',
        menutext  => 'BURST',
        module    => 'BURST',
        version   => '1.0.0',
        dbtype    => 'isolates,sequences',
        section   => 'postquery',
        order     => 10,
        system_flag => 'BURST',
        input     => 'query',
        requires  => 'mogrify',
        min       => 2,
        max       => 1000
    );
    return \%att;
}
```

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

```
BURST="yes"
```

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

5.9 Temporarily disabling database updates

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

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

5.9.1 Global

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

```
disable_updates=yes
```

An optional message can also be displayed by adding a `disable_update_message` value, e.g.

```
disable_update_message=The server is currently undergoing maintenance.
```

5.9.2 Database-specific

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

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

5.10 Host mapping

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

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

| | |
|----------------|-------------|
| #Existing_host | Mapped_host |
| server1 | server2 |
| localhost | server2 |

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

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

5.11 Improving performance

5.11.1 Use mod_perl

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

5.11.2 Cache scheme definitions within an isolate database

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

To do this use the update_scheme_caches.pl script found in the scripts/maintenance directory, e.g. to cache all schemes in the pubmlst_bigsdb_neisseria_isolates database

```
update_scheme_caches.pl -d pubmlst_bigsdb_neisseria_isolates
```

This script creates indexed tables within the isolate database called temp_scheme_X and temp_isolates_scheme_fields_1 (where X is the scheme_id). If these table aren't present, they are created as temporary tables every time a query is performed that requires a join against scheme definition data. This requires importing all profile definitions from the definitions database and determining scheme field values for all isolates. This may sound like it would be slow but caching only has a noticeable effect once you have >5000 profiles.

Note that you will need to run this script periodically as a CRON job to refresh the cache.

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

5.11.3 Use materialized views for scheme definitions

Because of the way BIGSdb allows any number of profile schemes to be set up, the data are stored in a normalised manner in multiple tables. A database view, e.g. `scheme_1`, is created that joins these tables so that they can be queried as you would a single table. A view, however, is only a pre-selected query rather than a physical table and you can not index columns on it to optimise query performance.

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

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

```
materialized_views="yes"
```

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

If you want an isolate database to benefit from this materialized view, make sure you put 'mv_scheme_X' (where X is the scheme id) in the `dbase_table` field (rather than 'scheme_X') when setting up the scheme in the isolate database configuration.

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

```
SELECT refresh_matview('mv_scheme_X');
```

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

5.11.4 Use a ramdisk for the secure temporary directory

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

5.12 Dataset partitioning

5.12.1 Sets

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

See also:

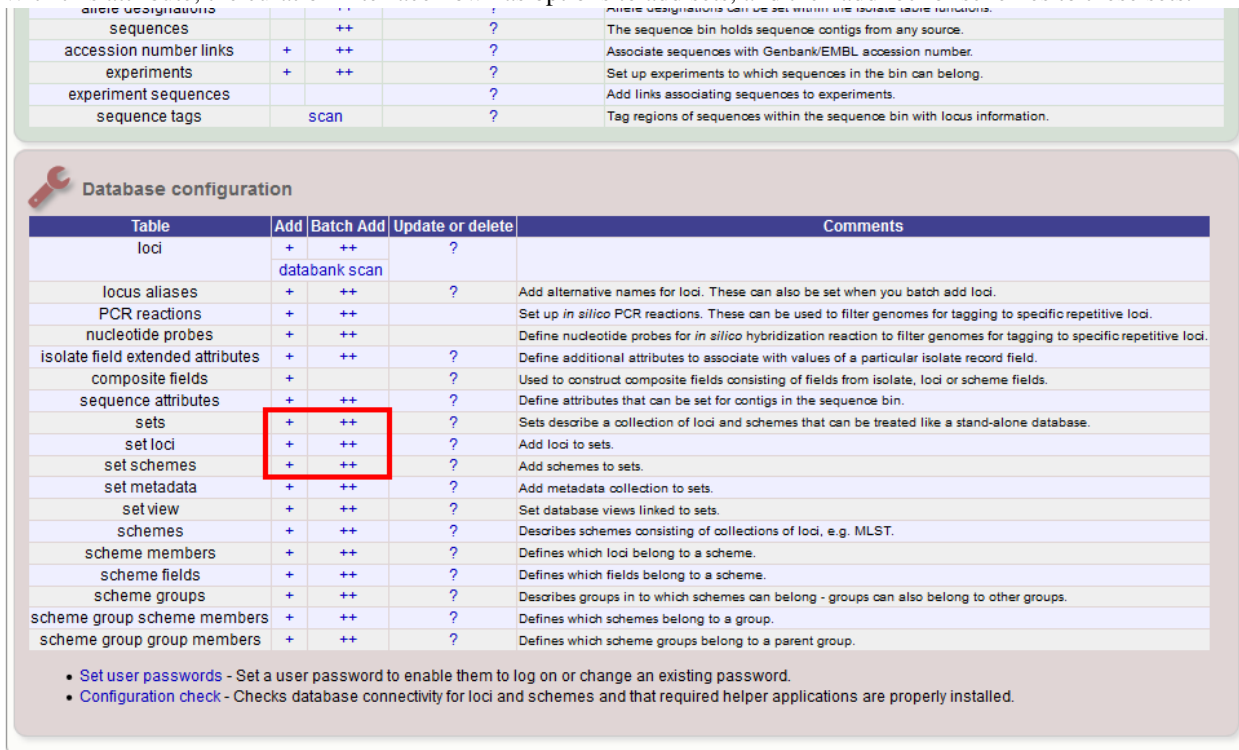
Sets (concept)

5.12.2 Configuration of sets

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

```
sets="yes"
```

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



The screenshot shows the 'Database configuration' interface. At the top, there is a small table with configuration options. Below it is a larger table with columns: Table, Add, Batch Add, Update or delete, and Comments. The 'sets' row is highlighted with a red box.


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

Below the table, there are two bullet points:

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


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

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


[Query: Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
[Breakdown: Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
[Links: Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Neisseria PubMLST database


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.



Datasets


This database contains multiple datasets. You can choose to display a single set or the whole database.

Please select: Whole database Choose




Query database

- [Search database](#) - advanced queries.
- [Browse database](#) - peruse all records.
- [Search by combinations of loci \(profiles\)](#) - including partial matching.
- [List query](#) - find isolates by matching a field to an entered list.




Option settings

- [Set general options](#) - including isolate table field handling.
- [Set display and query options for locus, schemes or scheme fields.](#)




General information

- Isolates: 20902
- Last updated: 2015-06-30
- [Update history](#)
- [About BIGSdb](#)




Breakdown

- [Single field](#)
- [Two field](#)
- [Unique combinations](#)
- [Scheme and alleles](#)
- [Publications](#)
- [Sequence bin](#)




Export

- [Export dataset](#)
- [Contigs](#)
- [Sequences](#) - XMFA / concatenated FASTA formats



Analysis

- [Codon usage](#)
- [Presence/absence status of loci](#)
- [Genome comparator](#)
- [BLAST](#)



Miscellaneous

- [Description of database fields](#)

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

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

```
set_id="1"
```

where the value is the name of the set.

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

5.12.3 Set metadata

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

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



```


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

```

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

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

| isolates | + | ++ | query browse list batch update | |
|---|---|------|--------------------------------------|--|
| isolate field extended attribute values | + | ++ | ? | Add values for additional isolate field attributes. |
| projects | + | ++ | ? | Set up projects to which isolates can belong. |
| project members | + | ++ | ? | Add isolates to projects. |
| isolate aliases | + | ++ | ? | Add alternative names for isolates. |
| PubMed links | + | ++ | ? | |
| allele designations | | ++ | ? | Allele designations can be set within the isolate table functions. |
| sequences | | ++ | ? | The sequence bin holds sequence contigs from any source. |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| experiments | + | ++ | ? | Set up experiments to which sequences in the bin can belong. |
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | | scan | ? | Tag regions of sequences within the sequence bin with locus information. |


Database configuration

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

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

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

```

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

GRANT SELECT,UPDATE,INSERT,DELETE ON meta_1 TO apache;

```

The above creates the database table for a metaset called ‘1’, defining new text fields for ‘town’ and ‘clinical_outcome’.

5.12.4 Set views

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

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

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

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

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

5.12.5 Using only defined sets

The only_sets attribute can be set to ‘yes’ to disable the option for ‘Whole database’ so that only sets can be viewed, e.g.

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

5.13 Adding new loci


See also:

Loci (concept)

5.13.1 Sequence definition databases

Single locus

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



Database configuration

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

- [Set user passwords](#) - Set a user password to enable them to log on or change an existing password.
- [Configuration check](#) - Checks database connectivity for loci and schemes and that required helper applications are properly installed.
- [Configuration repair](#) - Rebuild scheme tables

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

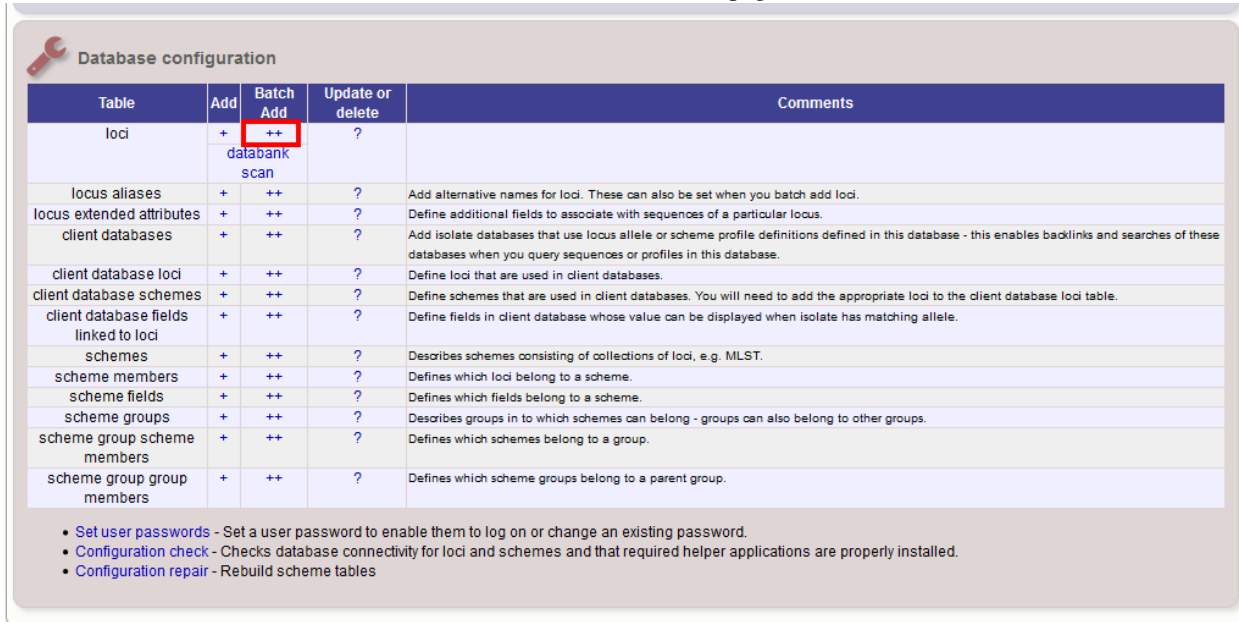
- `id` - The name of the locus.
 - Allowed: any value starting with a letter or underscore.
- `data_type` - Describes whether the locus is defined by nucleotide or peptide sequence.
 - Allowed: DNA/peptide.
- `allele_id_format` - The format for allele identifiers.
 - Allowed: integer/text.
- `length_varies` - Sets whether alleles can vary in length.
 - Allowed: true/false.
- `coding_sequence` - Sets whether the locus codes for a protein.
 - Allowed: true/false.
- `formatted_name` - Name with HTML formatting (optional).
 - This allows you to add formatting such as bold, italic, underline and superscripting to locus names as they appear in the web interface.
 - Allowed: valid HTML.
- `common_name` - The common name for the locus (optional).
 - Allowed: any value.
- `formatted_common_name` - Common name with HTML formatting (optional).
 - Allowed: valid HTML.
- `allele_id_regex` - [Regular expression](#) to enforce allele id naming (optional).
 - `^`: the beginning of the string
 - `$`: the end of the string
 - `d`: digit

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

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

Batch adding

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

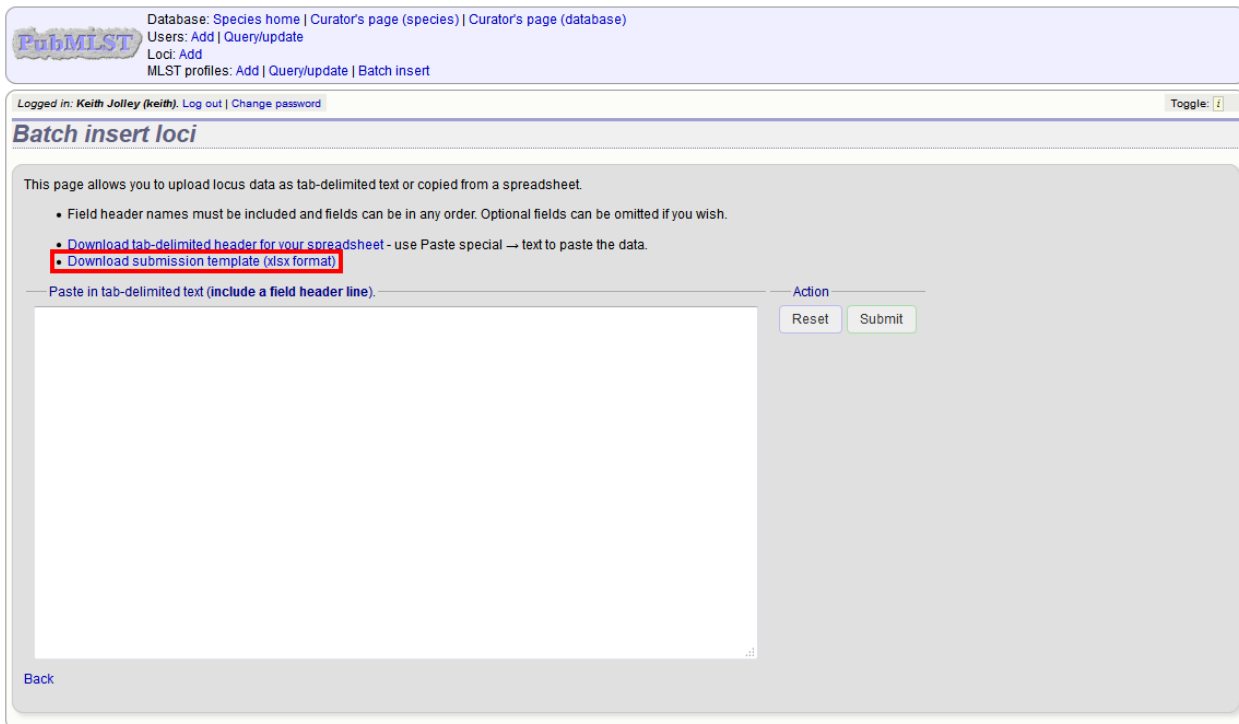


Database configuration

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

- [Set user passwords](#) - Set a user password to enable them to log on or change an existing password.
- [Configuration check](#) - Checks database connectivity for loci and schemes and that required helper applications are properly installed.
- [Configuration repair](#) - Rebuild scheme tables

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



Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Batch insert loci

This page allows you to upload locus data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)

Paste in tab-delimited text (include a field header line).

Action

[Back](#)


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

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

5.13.2 Isolate databases


Single locus

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



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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - Neisseria PubMLST


Add, update or delete records

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


Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-----------------------------------|-------------------------------|-----------|------------------|---|
| loci | + | ++ | ? | |
| | databank scan | | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attributes | + | ++ | ? | Define additional attributes to associate with values of a particular isolate record field. |
| composite fields | + | | ? | Used to construct composite fields consisting of fields from isolate, loci or scheme fields. |
| sequence attributes | + | ++ | | Define attributes that can be set for contigs in the sequence bin. |
| schemes | + | ++ | ? | Describe schemes consisting of collections of loci in a MLST |

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Add new locus

Please fill in the fields below - required fields are marked with an exclamation mark (!).

| Record | Action |
|------------------------|--|
| id:! | Reset Submit |
| data type:! | |
| allele id format:! | |
| length varies:! | |
| coding sequence:! | |
| curator:! | |
| date entered:! | |
| timestamp:! | |
| formatted name: | |
| common name: | |
| formatted common name: | |
| allele id regex: | |
| length: | |
| min length: | |
| max length: | |
| orf: | |
| genome position: | |
| match longest: | |
| complete cds: | |
| full name: | |
| product: | |
| description: | |

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

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

- This is useful for some loci that can have some sequences shorter than others, e.g. peptide loci defining antigenic loops. This can lead to instances of one sequence being longer than another but otherwise being identical. In these cases, usually the longer sequence is the one that should be matched.
 - Allowed: true/false.
- reference_sequence - Sequence used by the automated sequence comparison algorithms to identify sequences matching this locus. **This is only used if a sequence definition database has not been set up for this locus.**
- pcr_filter - Set to true if this locus is further defined by genome filtering using in silico PCR.
 - Allowed: true/false.
- probe_filter - Set to true if this locus is further defined by genome filtering using in silico hybridization.
 - Allowed: true/false.
- dbase_name - Name of database (system name).
 - Allowed: any text.
- dbase_host - Resolved name of IP address of database host - leave blank if running on the same machine as the isolate database.
 - Allowed: network address, e.g. 129.67.26.52 or zoo-oban.zoo.ox.ac.uk
- dbase_port - Network port on which the sequence definition database server is listening - leave blank unless using a non-standard port (5432).
 - Allowed: integer.
- dbase_user - Name of user with permission to access the sequence definition database - depending on the database configuration you may be able to leave this blank.
 - Allowed: any text (no spaces).
- dbase_password - Password of database user - again depending on the database configuration you may be able to leave this blank.
 - Allowed: any text (no spaces).
- dbase_table - Table in the sequence definition database that contains allele sequences for this locus. If the definition database uses BIGSdb this will be 'sequences'.
 - Allowed: any text (no spaces).
- dbase_id_field - Primary field in sequence database that defines allele. If the definition database uses BIGSdb this will be 'allele_id'.
 - Allowed: any text (no spaces).
- dbase_id2_field - Secondary field in sequence database that defines alleles. If dbase_id_field uniquely defines alleles for this locus then this should be left blank. If the definition database uses BIGSdb this will be 'locus'.
 - Allowed: any text (no spaces).
- dbase_id2_value - Secondary field value in sequence database that defines alleles. If dbase_id_field uniquely defines alleles for this locus then this should be left blank. If the definition database uses BIGSdb this will be the name of the locus used in the id field
 - Allowed: any text (no spaces).
- dbase_seq_field - Field in sequence database containing allele sequence. If the definition database uses BIGSdb this will be 'sequence'.
 - Allowed: any text (no spaces).

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

Using existing locus definition as a template

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

The screenshot shows the 'Add new locus' form in the BIGSdb interface. At the top, there is a navigation bar with links for 'Database: Species home | Curator's page (species) | Curator's page (database)', 'Users: Add | Query/update', and 'Isolates: Add | Query/update | Batch insert'. Below this, a status bar indicates 'Logged in: Keith Jolley (keith)' with links for 'Log out' and 'Change password'. The main heading is 'Add new locus', and a 'Show tools' link is visible in the top right corner, highlighted with a red box. The form itself is titled 'Please fill in the fields below - required fields are marked with an exclamation mark (!)'. It contains various fields for defining a new locus, including 'id', 'data type', 'allele id format', 'length varies', 'coding sequence', 'flag table', 'isolate display', 'main display', 'query field', 'analysis', 'curator', 'date entered', 'datestamp', 'formatted name', 'common name', 'formatted common name', 'allele id regex', 'length', 'orf', 'genome position', 'match longest', and 'reference sequence'. The 'Show tools' link is highlighted with a red box.

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: ?

Add new locus Hide tools

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

id:

data type:

allele id format: ?

length varies: ☐ true ☒ false ?

coding sequence: ☒ true ☐ false

flag table: ☒ true ☐ false ? Seqdef database supports allele flags

isolate display: ?

main display: ☐ true ☒ false ?

query field: ☒ true ☐ false ?

analysis: ☒ true ☐ false ?

curator: **Keith Jolley (keith)**

date entered: **2015-07-01**

timestamp: **2015-07-01**

formatted name: ?

common name: ?

formatted common name: ?

allele id regex: ?

length: ?

orf: ?

genome position: ?

match longest: ☐ true ☐ false ?

reference sequence:

Copy configuration from **NEIS0001 (lpxC)** Copy

All parameters will be copied except id, common name, reference sequence, genome position and length. The copied locus id will be substituted for 'PUT_LOCUS_NAME_HERE' in fields that include it.

The configuration will be copied over to the web form, with the exception of name fields. Some fields will require you to change the value 'PUT_LOCUS_NAME_HERE' with the value you enter in the id field. These are usually the dbase_id2_value, description_url and url fields:

allele id regex: ⓘ

length: ⓘ

orf: ⓘ

genome position: ⓘ

match longest: ☐ true ☐ false ⓘ

reference sequence:

pcr filter: ☐ true ☐ false ⓘ

probe filter: ☐ true ☐ false ⓘ

dbase name: ⓘ Name of the database holding allele sequences

dbase host: ⓘ IP address of database host

dbase port: ⓘ Network port accepting database connections

dbase user: ⓘ

dbase password: ⓘ

dbase table: ⓘ Database table that holds sequence information for this locus

dbase id field: ⓘ Primary field in sequence database that defines allele, e.g. 'allele_id'

dbase id2 field: ⓘ Secondary field that defines allele, e.g. 'locus'

dbase id2 value: ⓘ Secondary field value, e.g. locus name

dbase seq field: ⓘ Field in sequence database containing allele sequence

description url: ⓘ

url: ⓘ

submission template: ☐ true ☒ false ⓘ Include column in isolate submission template for this locus


aliases:

Action

Complete the form and click 'Submit'.


Batch adding

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



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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - Neisseria PubMLST


Add, update or delete records

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


Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-----------------------------------|-----|-------------------------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attributes | + | ++ | ? | Define additional attributes to associate with values of a particular isolate record field. |
| composite fields | + | | ? | Used to construct composite fields consisting of fields from isolate, loci or scheme fields. |

Click the link to download an Excel template:

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [2](#)

Batch insert loci

This page allows you to upload locus data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- Enter aliases (alternative names) for your locus as a semi-colon (;) separated list.
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)

Paste in tab-delimited text (include a field header line).

Action

[Back](#)


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

5.14 Defining locus extended attributes

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

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

| | | | | |
|-----------------------------|---|------------|--------------------------------------|---|
| user groups | + | ++ | | Users can be members of these groups - use for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| MLST profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | ? | |


Database configuration

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

- [Set user passwords](#) - Set a user password to enable them to log on or change an existing password.
- [Configuration check](#) - Checks database connectivity for loci and schemes and that required helper applications are properly installed.
- [Configuration repair](#) - Rebuild scheme tables

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Add new locus extended attribute

Please fill in the fields below - required fields are marked with an exclamation mark (!).

| Record | Action |
|---|--|
| locus: <input type="text"/> | <input type="button" value="Reset"/> <input type="button" value="Submit"/> |
| field: <input type="text"/> | |
| value format: <input type="text" value="text"/> | |
| required: <input type="radio"/> true <input checked="" type="radio"/> false ? | |
| curator: Keith Jolley (keith) | |
| datestamp: 2015-07-01 | |
| value regex: <input type="text"/> ? | |
| description: <div style="border: 1px solid #ccc; height: 100px; width: 100%;"></div> | |
| option list: <div style="border: 1px solid #ccc; height: 100px; width: 100%;"></div> | |
| length: <input type="text"/> | |
| field order: <input type="text"/> | |

- 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.

5.15 Defining schemes

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

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

To set up a new scheme, you need to:

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

See also:


Schemes (concept)

5.15.1 Sequence definition databases

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

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

| | | | | |
|----------------------------|---|----|---|--|
| sequences) | | | | |
| MLST profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | ? | |



Database configuration

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

- [Set user passwords](#) - Set a user password to enable them to log on or change an existing password.
- [Configuration check](#) - Checks database connectivity for loci and schemes and that required helper applications are properly installed.
- [Configuration repair](#) - Rebuild scheme tables

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

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


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)
Toggle: [?](#)

Add new scheme

Please fill in the fields below - required fields are marked with an exclamation mark (!).


Record

id: 1
description: MLST
curator: Keith Jolley (keith)
datestamp: 2014-07-04
date entered: 2014-07-04
display order:
allow missing loci: ☐ true ☐ false

Action
Reset Submit

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


| | | | | |
|----------------------------|---|----|--------------------------------------|--|
| sequences) | | | | |
| MLST profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | ? | |


Database configuration

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

- [Set user passwords](#) - Set a user password to enable them to log on or change an existing password.
- [Configuration check](#) - Checks database connectivity for loci and schemes and that required helper applications are properly installed.
- [Configuration repair](#) - Rebuild scheme tables

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


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)
Toggle: ?

Add new scheme 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 affect allele designations, but any profiles will have to be reloaded.

Please fill in the fields below - required fields are marked with an exclamation mark (!).

| Record | Action |
|--|--------------|
| scheme id: MLST locus: abcZ curator: Keith Jolley (keith) timestamp: 2014-07-04 field order: 1 | Reset Submit |

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


| | | | | |
|----------------------------|---|----|--------------------------------------|--|
| sequences) | | | | |
| MLST profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | ? | |


Database configuration

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

- [Set user passwords](#) - Set a user password to enable them to log on or change an existing password.
- [Configuration check](#) - Checks database connectivity for loci and schemes and that required helper applications are properly installed.
- [Configuration repair](#) - Rebuild scheme tables

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


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)
Toggle: ?

Add new scheme field

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 affect allele designations, but any profiles will have to be reloaded.

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

Action

scheme id: MLST

field: ST

type: integer

primary key: ☒ true ☐ false !

dropdown: ☐ true ☒ false !

curator: Keith Jolley (keith)

timestamp: 2014-07-04

description:

field order: 1

index: ☐ true ☐ false !

value regex:

Reset Submit

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


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

5.15.2 Isolate databases

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

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


| Table | Add | Batch Add | Update or delete | Comments |
|------------------------|-----|-----------|------------------|--|
| sequences | + | ++ | ? | The sequence bin stores sequence contigs from any source. |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| experiments | + | ++ | ? | Set up experiments to which sequences in the bin can belong. |
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | | scan | ? | Tag regions of sequences within the sequence bin with locus information. |



Database configuration

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

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

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


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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)
Toggle: 

Add new scheme

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

id: 41
description:!
isolate display: ☒ true ☐ false
main display: ☒ true ☐ false
query field: ☒ true ☐ false
query status: ☐ true ☒ false
analysis: ☒ true ☐ false
curator: Keith Jolley (keith)
timestamp: 2015-07-01
date entered: 2015-07-01
dbase name:
dbase host:
dbase port:
dbase user:
dbase password:
dbase table:
display order:
allow missing loci: ☐ true ☐ false

Action
Reset Submit


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

- Allowed: any positive integer.
- description - Short description - this is used in tables so make sure it's not too long.
 - Allowed: any text.
- isolate_display - Sets whether or not fields for this scheme are displayed in a detailed isolate record - this can be overridden by user preference.
 - Allowed: allele only/sequence/hide.
- main_display - Sets whether or not fields for this scheme are displayed in a main results table by default - this can be overridden by user preference.
 - Allowed: true/false.
- query_field - Sets whether or not fields for this scheme can be used in queries by default - this can be overridden by user preference.
 - Allowed: true/false.
- query_status - Sets whether a dropdown list box should be displayed in the query interface to filter results based on profile completion for this scheme - this can be overridden by user preference.
 - Allowed: true/false.
- analysis - Sets whether or not alleles for this locus can be used in analysis functions by default - this can be overridden by user preference.
 - Allowed: true/false.
- dbase_name - Name of seqdef database (system name) containing scheme profiles (optional).
 - Allowed: any text.
- dbase_host - Resolved name of IP address of database host - leave blank if running on the same machine as the isolate database (optional).
 - Allowed: network address, e.g. 129.67.26.52 or zoo-oban.zoo.ox.ac.uk
- dbase_port - Network port on which the sequence definition database server is listening - leave blank unless using a non-standard port, 5432 (optional).
 - Allowed: integer.
- dbase_user - Name of user with permission to access the sequence definition database - depending on the database configuration you may be able to leave this blank (optional).
 - Allowed: any text (no spaces).
- dbase_password - Password of database user - again depending on the database configuration you may be able to leave this blank (optional).
 - Allowed: any text (no spaces).
- dbase_table - Table in the sequence definition database that contains profiles for this scheme. If the definition database uses BIGSdb this will be 'scheme_X' where X is the scheme id number in the seqdef database.
 - Allowed: any text (no spaces).
- display_order - Integer reflecting the display position for this scheme within the interface (optional).
 - Allowed: any integer.
- allow_missing_loci - Allow profile definitions to contain 'O' (locus missing) or 'N' (any allele).

5.16 Organizing schemes into hierarchical groups

Schemes can be organized in to groups, and these groups can in turn be members of other groups. This facilitates 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.


Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FeTA](#) | [Options](#) | [Isolate Database](#)

[Help](#)

Download allele sequences

Select loci by scheme | [Alphabetical list](#) | [All loci by scheme](#)

Click within the tree to display details of loci belonging to schemes or groups of schemes - clicking a group folder will display the loci for all schemes within the group and any subgroups. Click the nodes to expand/collapse.

- All loci
 - Capsule
 - Genetic Information Processing
 - Replication and Repair
 - DNA replication
 - Nucleotide excision repair
 - Transcription
 - Translation
 - Metabolism
 - Typing
 - MLST
 - Finotyping antigens
 - Antigen genes
 - eMLST (20 locus partial genes)
 - eMLST (20 locus whole genes)
 - Other schemes
 - Loci not in schemes


MLST

| Locus | Download | Type | Alleles | Length | Full name/product | Curator(s) | Last updated |
|----------------------|--------------------------|------|---------|------------------------------|-------------------------------|--|--------------|
| adk | Download | DNA | 426 | Fixed: 465 bp | test_full_name / test_product | O. Harrison, K. Jolley | 2015-01-15 |
| abcZ | Download | DNA | 630 | Variable: (432 min; 433 max) | | O. Harrison, K. Jolley | 2015-04-14 |
| aroE | Download | DNA | 682 | Fixed: 490 bp | | O. Harrison, K. Jolley | 2013-01-22 |
| fumC | Download | DNA | 639 | Fixed: 465 bp | | O. Harrison | 2013-01-23 |
| gdh | Download | DNA | 653 | Fixed: 501 bp | | O. Harrison | 2013-01-22 |
| pdhC | Download | DNA | 652 | Fixed: 480 bp | | O. Harrison | 2013-01-21 |
| pgm | Download | DNA | 675 | Fixed: 450 bp | | O. Harrison | 2013-01-22 |

Download table: [tab-delimited text](#) | [Excel format](#)

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

| | | | | |
|------------------------|---|------|---|--|
| projects | + | ++ | ? | Set up projects to which isolates can belong. |
| project members | + | ++ | ? | Add isolates to projects. |
| isolate aliases | + | ++ | ? | Add alternative names for isolates. |
| PubMed links | + | ++ | ? | |
| allele designations | | ++ | ? | Allele designations can be set within the isolate table functions. |
| sequences | | ++ | ? | The sequence bin holds sequence contigs from any source. |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| experiments | + | ++ | ? | Set up experiments to which sequences in the bin can belong. |
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | | scan | ? | Tag regions of sequences within the sequence bin with locus information. |


Database configuration

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

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

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Add new scheme group

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

id: 1
 name: !
 curator: Keith Jolley (keith)
 timestamp: 2015-07-01
 description:
 display order:
 seq query: ☒ true ☐ false !

Action

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

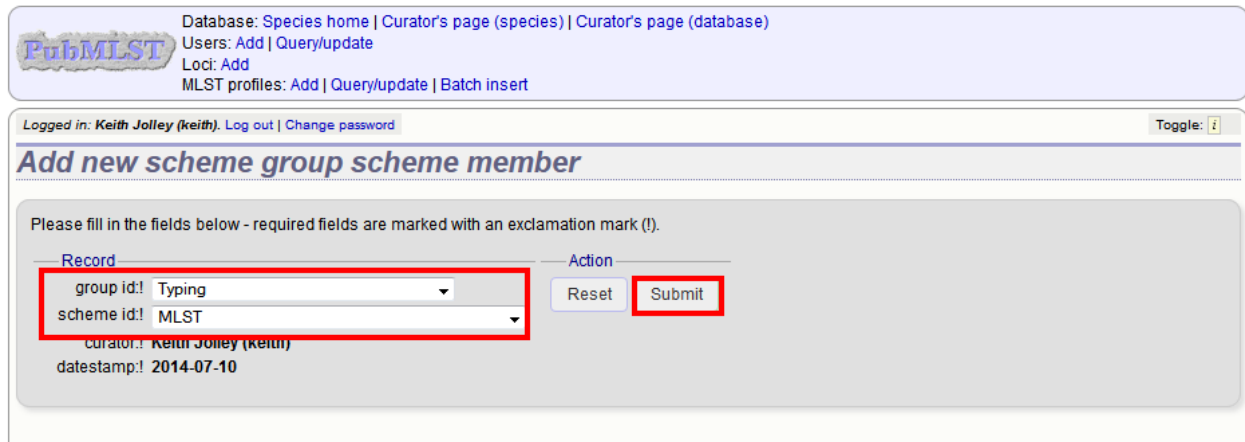
Schemes can be added to groups by clicking the add (+) scheme group scheme members link.

Database configuration

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

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

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



Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loc: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle:

Add new scheme group scheme member

Please fill in the fields below - required fields are marked with an exclamation mark (!).

| Record | Action |
|--|--|
| group id!: <input type="text" value="Typing"/> | <input type="button" value="Reset"/> <input type="button" value="Submit"/> |
| scheme id!: <input type="text" value="MLST"/> | |
| curator!: Keith Jolley (keith) | |
| datestamp!: 2014-07-10 | |

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


5.17 Setting up client databases

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

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

Multiple client databases can be queried simultaneously.

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


 **Database configuration**

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

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

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: **Keith Jolley (keith)**. [Log out](#) | [Change password](#) Toggle: 

Add new client database

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record **Action**
[Reset](#) [Submit](#)

id:

name:

description:

dbase name: Name of the database holding isolate data

dbase config name: Name of the database configuration

curator:

datestamp:

dbase host: IP address of database host

dbase port: Network port accepting database connections

dbase user:

dbase password:

dbase view: View of isolates table to use


url: Web URL to database script

- **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.

5.17.1 Look up isolates with given allele

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

 **Database configuration**

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

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

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.

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#) Toggle: [?]

Add new locus to client database definition

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record Action

client dbase id: 1) PubMLST isolates Reset Submit

locus: abcZ

curator: Keith Jolley (keith)

datestamp: 2014-07-04

locus alias: name that this locus is referred by in client database (if different)

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FeTA | Options | Isolate Database

Allele information - *abcZ*: 5

Provenance/meta data

locus: *abcZ*
 allele: 5

sequences: TTTGATACCG TTGCCGAAAG TTGGGCGAA ATTGCGGATT TATTGCGCCG TTATCATCAT GTACGCCATG ASITGGAAAA TGGTTCGAGT GAGGCTTTGT TGAAGAGCT TAACGAATTG
 CAACTTGAAG TCGAAGCGAA GGACGGCTGG AAGCTGGATG CGGCAGTCAA GCAGACTTTG GGTGAACCTG GTTTCGCCAG AACGAAAAA ATCGGCAACC TCTCGGCGGG ACAGAAAAAG
 CGTGTTCGCC TAGCGCAGGC TTGGGTGCAG AAGCTGATG TATTGCTGCT GGACGAACCG ACCAACCATT TGGACATTGA CGCGATTATT TGGCTGGAAA ATCTGCTTAA AGCGTTTGAA
 GGCAGCCTGG TTGTGATTAC CCACGACCGC CGTTTTTTGG ACAATATGCG CACGCGCATC GTCGAATCG ATC

length: 433
 status: Sanger trace checked
 date entered: 2001-02-07
 datestamp: 2009-11-11
 sender: Keith Jolley, University of Oxford, UK
 curator: Man-Suen Chan, University of Oxford (E-mail: man-suen.chan@paediatrics.ox.ac.uk)

Profiles containing this allele

MLST: 183 profiles

Isolate databases

PubMLST isolates: Contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.
 399 isolates

5.17.2 Look up isolates with a given scheme primary key

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

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FeTA | Options | Isolate Database

Profile information for ST-11 (MLST)

| ST | <i>abcZ</i> | <i>adk</i> | <i>aroE</i> | <i>fumC</i> | <i>gdh</i> | <i>pdhC</i> | <i>pgm</i> | clonal complex |
|----|-------------|------------|-------------|-------------|------------|-------------|------------|-----------------------------|
| 11 | 2 | 3 | 4 | 3 | 8 | 4 | 6 | ST-11 complex/ET-37 complex |

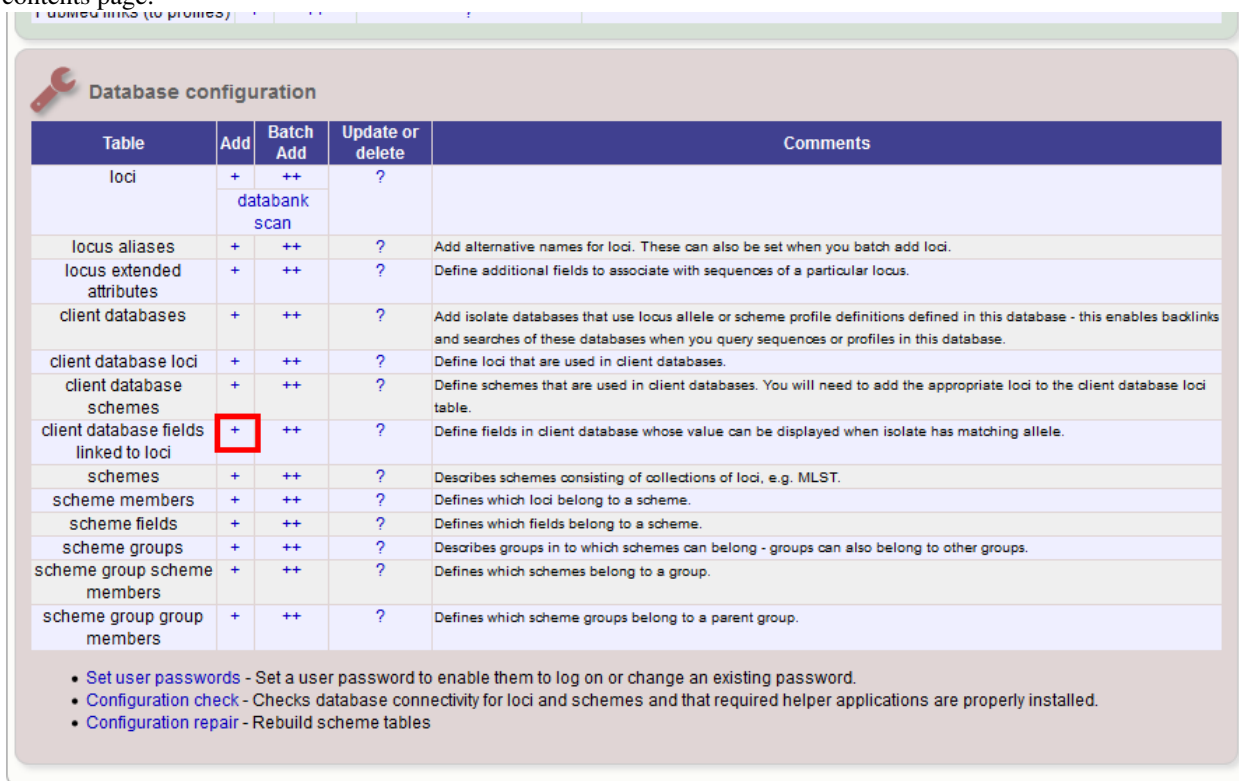
sender: Paula Kriz, Paula Kriz and Keith Jolley
 curator: Keith Jolley, University of Oxford, UK (E-mail: keith.jolley@zoo.ox.ac.uk)
 date entered: 2001-02-07
 datestamp: 2009-11-11

Client database

PubMLST isolates: Contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.
 2288 isolates

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

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

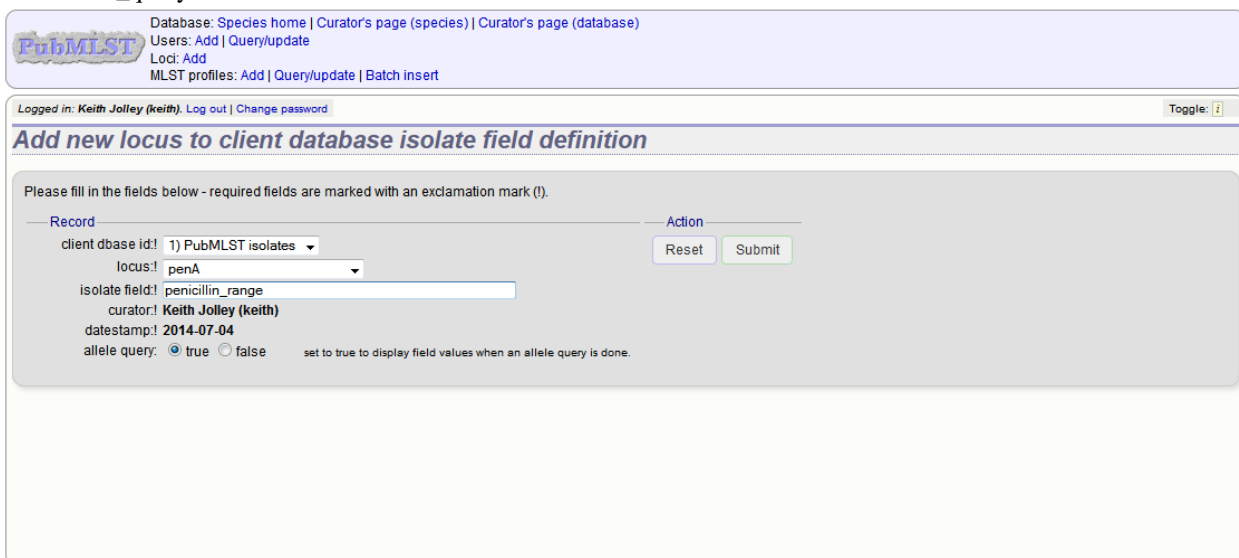


Database configuration

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

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

Select the client database and locus from the dropdown lists and enter the isolate database field that you’d like to link. The ‘allele_query’ field should be set to true.



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

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

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: 2014-07-04

allele query: ☒ true ☐ false set to true to display field values when an allele query is done.

Action

Reset Submit

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.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FeTA](#) | [Options](#) | [Isolate Database](#)

Allele information - penA: 9

Provenance/meta data

locus: penA
allele: 9

sequences: GACGGCGTIT TGCTGCCGGT CAGCTTTGAA AACAGGCGG TTGGGCCGCA AGGCAACGT ATATTITAAAG CATCGACCGC ACCTCAGGTG CGTGAGTTGA TGGTTTCTGT AACGGAACCT
GGCGGTACGG GTACGGCGGG TGGGTAGAT GGTTCGACG TCGGGCGAAA AACCGGTACG GGGGTAAAGT TGGTTAACGG TGGTTACGTC GATTACAAAC ACCTTGCCAC TTTCATCGGT
TTTGCCCGGG CTAAAAATCC GCGTGTGATT GTGGCGGTAA CCATTGACGA GCCGACTGCA AACGGTTACT ACGGCGCGCT AGTGACAGGT CCGGTCTTCA AACAAATTAT GGGCGGTAGC
CTGAACATCT TGGCGTTTC TCGACCAA 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 (E-mail: keith.jolley@zoo.ox.ac.uk)

mutation F504L: yes
mutation A510V: yes
mutation I515V: yes
mutation H541N: yes
mutation I566V: yes

Publication (1)

- Taha MK, Vázquez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, Cafferkey MT, Carion F, Christensen JJ, Diggle M, Edwards G, Enríquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lambertsen L, Levenet I, Musilek M, Paragi M, Sagner A, Skoczynska 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

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.

141 isolates

Linked data

penicillin_range: >0.06 - 1 (intermediate) [n=86] [PubMLST isolates](#)

5.18 Rule-based sequence queries

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

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

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

5.18.1 Example rule code

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

The code for this rule is as follows:

```
#Clinical identification rule

#Update job viewer status
update_status({stage=>'Scanning MLST loci'});

#Scan genome against all scheme 1 (MLST) loci
scan_scheme(1);
```

```

#Update job viewer status
update_status({percent_complete=>30, stage=>'Scanning PorA and FetA VRs'});

#Scan genome against the PorA VR and FetA VR loci
scan_locus($_) foreach qw(PorA_VR1 PorA_VR2 FetA_VR);

Add text to main output
append_html("<h1>Strain type</h1>");

#Set variables for the scanned results. These can be found in the
#$results->{'locus'} hashref
my %alleles;
$alleles{$_} = $results->{'locus'}->{$_} // 'ND' foreach qw(PorA_VR1 PorA_VR2);
$alleles{'FetA_VR'} = $results->{'locus'}->{'FetA_VR'} // 'F-ND';

#Scheme field values are automatically determined if a complete
#profile is available. These are stored in the $results->{'scheme'} hashref
my $st = $results->{'scheme'}->{1}->{'ST'} // 'ND';
append_html("<ul><li>P1.$alleles{'PorA_VR1'}, $alleles{'PorA_VR2'}; $alleles{'FetA_VR'}; ST-$st ");

#Reformat clonal complex using a regular expression, e.g.
#'ST-11 clonal complex/ET-37 complex' gets rewritten to 'cc11'
my $cc = $results->{'scheme'}->{1}->{'clonal_complex'} // '-';
$cc =~ s/ST-(\S+) complex.*cc$1/;

append_html("$cc</li></ul>");
if ($st eq 'ND'){
    append_html("<p>ST not defined. If individual MLST loci have been found "
        . "they will be displayed below:</p>");

    #The get_scheme_html function automatically formats output for a scheme.
    #Select whether to display in a table rather than a list, list all loci, and/or list fields.
    append_html(get_scheme_html(1, {table=>1, loci=>1, fields=>0}));
}

#Antibiotic resistance
update_status({percent_complete=>80, stage=>'Scanning penA and rpoB'});
scan_locus($_) foreach qw(penA rpoB);
if (defined $results->{'locus'}->{'penA'} || defined $results->{'locus'}->{'rpoB'}) {
    append_html("<h1>Antibiotic resistance</h1><ul>");
    if (defined $results->{'locus'}->{'penA'}) {
        append_html("<li><i>penA</i> allele: $results->{'locus'}->{'penA'}");

        #If a client isolate database has been defined and values have been defined in
        #the client_dbase_loci_fields table, the values for a field in the isolate database can be
        #retrieved based on isolates that have a particular allele designated.
        #The min_percentage attribute states that only values that are represented by at least that
        #proportion of all isolates that had a value set are returned (null values are ignored).
        my $range = get_client_field(1, 'penA', 'penicillin_range', {min_percentage => 75});
        append_html(" (penicillin MIC: $range->[0]->{'penicillin_range'})") if @$range;
        append_html("</li>");
    }
    if (defined $results->{'locus'}->{'rpoB'}) {
        append_html("<li><i>rpoB</i> allele: $results->{'locus'}->{'rpoB'}");
        my $range = get_client_field(1, 'rpoB', 'rifampicin_range', {min_percentage => 75});
        append_html(" (rifampicin MIC: $range->[0]->{'rifampicin_range'})") if @$range;
        append_html("</li>");
    }
}

```

```
append_html("</ul>");
}
```

Rule files

The rule file is placed in a rules directory within the database configuration directory, e.g. `/etc/bigsdb/dbase/pubmlst_neisseri_seqdef/rules`. Rule files are suffixed with `‘.rule’` and their name should be descriptive since it is used within the interface, i.e. the above rule file is named `Clinical_identification.rule` (underscores are converted to spaces in the web interface).

Linking to the rule query

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

http://pubmlst.org/perl/bigsdb/bigsdb.pl?db=pubmlst_neisseria_seqdef&page=plugin&name=RuleQuery&ruleset=Clinical_identification

To place a link to this within the database contents page an HTML file called `job_query.html` can be placed in a contents directory within the database configuration directory, e.g. in `/etc/bigsdb/dbases/pubmlst_neisseria_seqdef/contents/job_query.html`. This file should contain a list entry (i.e. surrounded with `` and `` tags) that will appear in the ‘Query database’ section of the contents page.

Adding descriptive text

Descriptive text for the rule, which will appear on the rule query page, can be placed in a file called `description.html` in a directory with the same name as the rule within the rule directory, e.g. in `/etc/bigsdb/dbases/pubmlst_neisseria_seqdef/rules/Clinical_identification/description.html`.

5.19 Workflow for setting up a MLST scheme

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

Seqdef database

1. Create appropriate loci
2. Create new scheme ‘MLST’
3. Add scheme_field ‘ST’ with `primary_key=TRUE` (add `clonal_complex` if you want; set this with `primary_key=FALSE`)
4. Add each locus as a scheme_member
5. You’ll then be able to add profiles

Isolate database

1. Create the same loci with the following additional parameters (example locus ‘atpD’)
 - `dbase_name`: `seqdef_db`
 - `dbase_table`: `sequences`
 - `dbase_id_field`: `allele_id`
 - `dbase_id2_field`: `locus`


- dbase_id_value: atpD
 - dbase_seq_field: sequence
 - url: something like /cgi-bin/bigsdb/bigsdb.pl?db=seqdef_db&page=alleleInfo&locus=atpD&allele_id=[?]
- Create scheme 'MLST' with:
 - dbase_name: seqdef_db
 - dbase_table: scheme_1 (or whatever the id of your seqdef scheme is)
 - Add scheme_field ST as before
 - Add loci as scheme_members

5.20 Defining new loci based on annotated reference genome

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

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

| | | | | |
|------------------------|---|------|---|--|
| allele designations | | ++ | ? | Allele designations can be set within the isolate table functions. |
| sequences | | ++ | ? | The sequence bin holds sequence contigs from any source. |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| experiments | + | ++ | ? | Set up experiments to which sequences in the bin can belong. |
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | | scan | ? | Tag regions of sequences within the sequence bin with locus information. |


Database configuration

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

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

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Scan EMBL/Genbank record for loci

This function allows you to scan an EMBL or Genbank (whole genome) file in order to create a batch upload file for setting up new loci.

— Please enter accession number — Primary identifier — Action —

Accession: ☒ locus tag ☐ gene name

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Scan EMBL/Genbank record for loci

This function allows you to scan an EMBL or Genbank (whole genome) file in order to create a batch upload file for setting up new loci.

— Please enter accession number — Primary identifier — Action —

Accession: ☒ locus tag ☐ gene name

Download table: [tab-delimited text](#) | [Excel format](#) (suitable for batch upload of loci).

Download alleles: [tab-delimited text](#) | [Excel format](#) (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 | Length |
|---------|---------------|---|--------|
| NMC0001 | lpxC; envA | UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase | 924 |
| NMC0002 | pilS1 | pilin (fragment) | 291 |
| NMC0003 | pilS2 | truncated pilin | 366 |
| NMC0004 | fbp | peptidyl-prolyl cis-trans isomerase | 330 |
| NMC0005 | | putative membrane protein | 219 |
| NMC0006 | | putative glycerate dehydrogenase | 954 |
| NMC0007 | metG | methionyl-tRNA synthetase | 2058 |
| NMC0008 | glmS | glucosamine-fructose-6-phosphate aminotransferase [isomerizing] | 1839 |
| NMC0009 | | putative lipoprotein | 519 |
| NMC0010 | gna33 | outer membrane lipoprotein Gna33 [1] | 1326 |
| NMC0011 | | putative integral membrane protein | 840 |
| NMC0012 | | putative lipoprotein | 1167 |
| NMC0013 | | possible membrane protein | 1266 |
| NMC0014 | phnA | putative phosphonoacetate hydrolase | 330 |
| NMC0015 | glmU | bifunctional GlmU protein [includes: UDP-N-acetylglucosamine pyrophosphorylase (EC 2.7.7.23) (N-acetylglucosamine-1-phosphate uridylyltransferase); | 1371 |

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.


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Scan EMBL/Genbank record for loci

This function allows you to scan an EMBL or Genbank (whole genome) file in order to create a batch upload file for setting up new loci.

Please enter accession number Primary identifier Action

Accession:
☒ locus tag

☐ gene name

Download table: [tab-delimited text](#) | [Excel format](#) (suitable for batch upload of loci).
Download alleles: [tab-delimited text](#) | [Excel format](#) (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 | Length |
|---------|---------------|---|--------|
| NMC0001 | lpxC; envA | UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase | 924 |
| NMC0002 | pilS1 | pilin (fragment) | 291 |
| NMC0003 | pilS2 | truncated pilin | 366 |
| NMC0004 | fbp | peptidyl-prolyl cis-trans isomerase | 330 |
| NMC0005 | | putative membrane protein | 219 |
| NMC0006 | | putative glycerate dehydrogenase | 954 |
| NMC0007 | metG | methionyl-tRNA synthetase | 2058 |
| NMC0008 | glmS | glucosamine-fructose-6-phosphate aminotransferase [isomerizing] | 1839 |
| NMC0009 | | putative lipoprotein | 519 |
| NMC0010 | gna33 | outer membrane lipoprotein Gna33 [1] | 1326 |
| NMC0011 | | putative integral membrane protein | 840 |
| NMC0012 | | putative lipoprotein | 1167 |
| NMC0013 | | possible membrane protein | 1266 |
| NMC0014 | phnA | putative phosphonoacetate hydrolase | 330 |
| NMC0015 | glmU | bifunctional GlmU protein [includes: UDP-N-acetylglucosamine pyrophosphorylase (EC 2.7.7.23) (N-acetylglucosamine-1-phosphate uridylyltransferase); | 1371 |

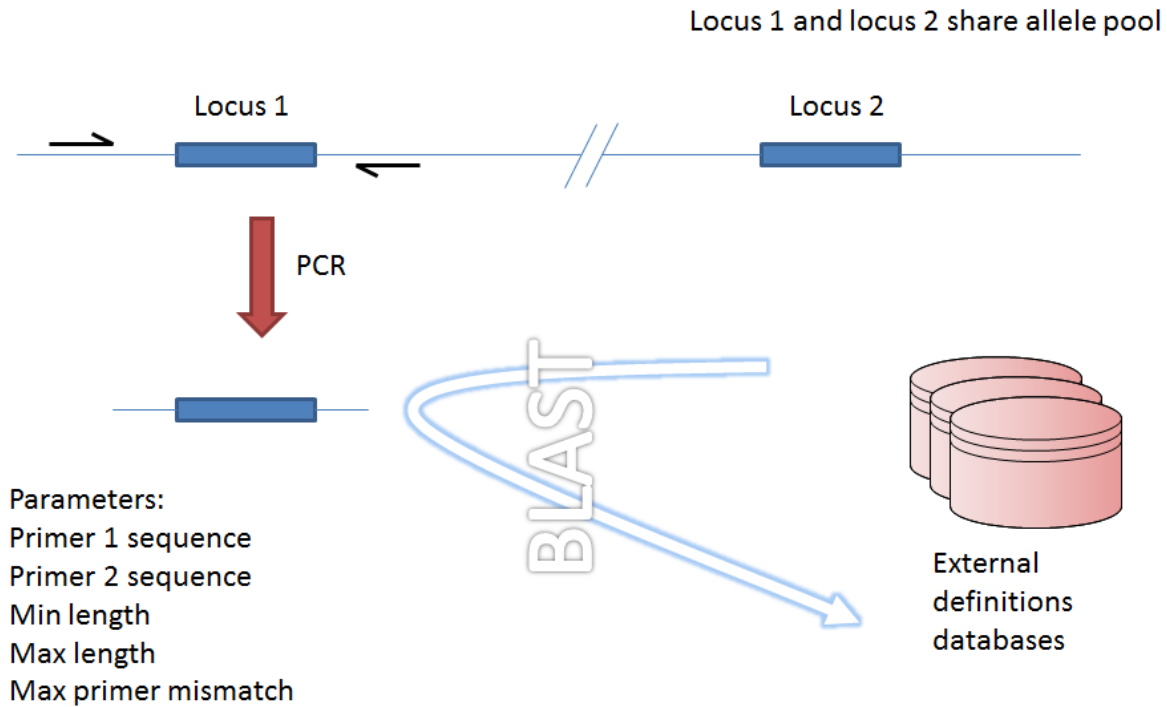
5.21 Genome filtering

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

5.21.1 Filtering by *in silico* PCR

Provided a locus can be predicted to be specifically amplified 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.

Fig. 5.1: Genome filtering by *in silico* PCR.

| | | | | |
|------------------------|---|------|---|--|
| allele designations | | ++ | ? | Allele designations can be set within the isolate table functions. |
| sequences | | ++ | ? | The sequence bin holds sequence contigs from any source. |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| experiments | + | ++ | ? | Set up experiments to which sequences in the bin can belong. |
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | | scan | ? | Tag regions of sequences within the sequence bin with locus information. |

Database configuration

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

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

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.

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle:

Add new PCR reaction

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

id:

description:

primer1:

primer2:

curator: **Keith Jolley (keith)**

datestamp: **2015-07-01**

min length: Minimum length of product to return

max length: Maximum length of product to return


max primer mismatch: Maximum sequence mismatch per primer

Action

- 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.

| | | | | |
|------------------------|---|------|---|--|
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| experiments | + | ++ | ? | Set up experiments to which sequences in the bin can belong. |
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | | scan | ? | Tag regions of sequences within the sequence bin with locus information. |


Database configuration

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

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

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

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

5.21.2 Filtering by *in silico* hybridization

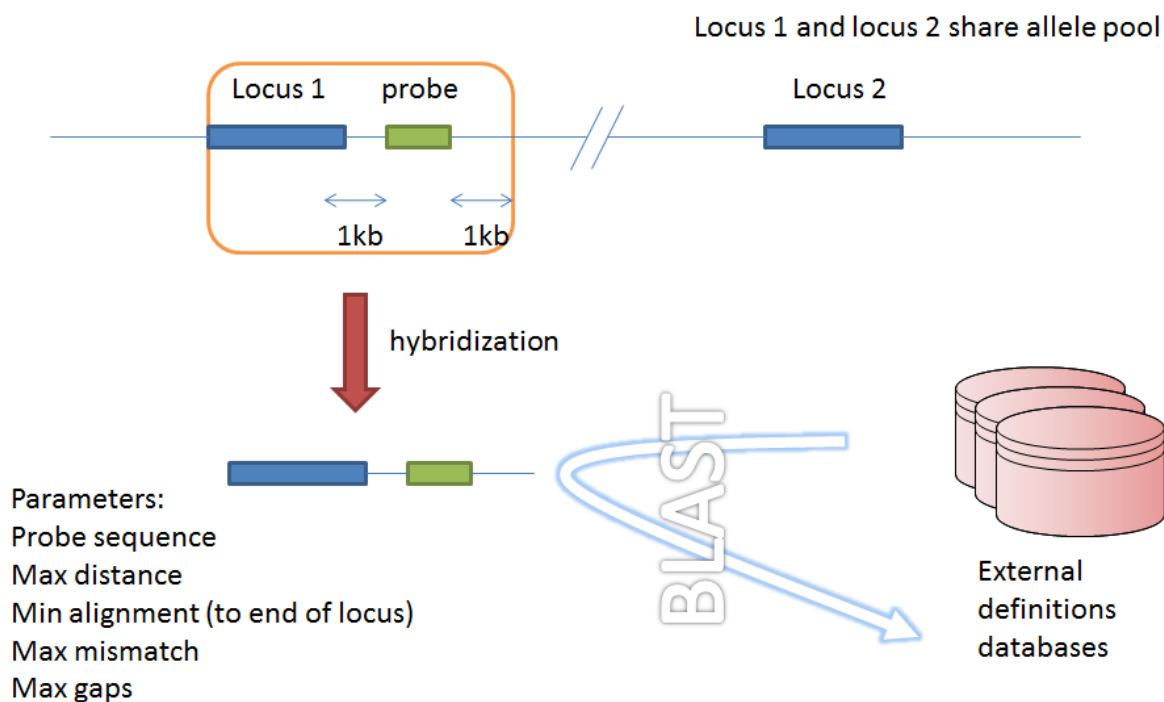
An alternative is to define a locus by proximity to a single probe sequence. This is especially useful if you have multiple contigs and the locus in question may be at the end of a contig so that it doesn't have upstream or downstream sequence available for PCR filtering.

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

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

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

- `probe_id` - Dropdown list of probe names.
 - Allowed: selection from list.
- `locus` - Dropdown list of loci.
 - Allowed: selection from list.
- `max_distance` - Minimum distance of probe from end of locus.

Fig. 5.2: Filtering by *in silico* hybridization

- Allowed: any positive integer.
- min_alignment - Minimum length of alignment allowed.
 - Allowed: any positive integer.
- max_mismatch - Maximum number of mismatches allowed in alignment.
 - Allowed: any positive integer.
- max_gaps - Maximum number of gaps allowed in alignment.
 - Allowed: any positive integer.

Finally edit the locus table and set the probe_filter field for the specified locus to 'true'.


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

5.22 Setting locus genome positions

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

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


| | | | | |
|------------------------|---|------|---|--|
| sequence | | | | The sequence bin holds sequences belonging to any locus. |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| experiments | + | ++ | ? | Set up experiments to which sequences in the bin can belong. |
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | | scan | ? | Tag regions of sequences within the sequence bin with locus information. |


Database configuration

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

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

| | | | | |
|----------------------|---|------|---|--|
| experiments | + | ++ | ? | Set up experiments to which sequences in the bin can belong. |
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | | scan | ? | Tag regions of sequences within the sequence bin with locus information. |


Database configuration


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

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* PubMLST

 Add, update or delete records

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

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [i](#) Field help: [id](#) [Go](#)

Isolate query/update

Isolate provenance/phenotype fields: isolate = MC58 Display/sort options: Order by: id ascending Display: 25 records per page

Action:

1 record returned. Click the hyperlink for detailed information.

Delete Tag scanning Projects

| Delete | Update | Sequence bin | New version | Isolate fields | | | | species | serogroup | ST | MLST | Finotyping antigens | | | test | test3 | Loci | | |
|--------|--------|--------------|-------------|---------------------|---------|---------|---------|---------|-----------|------------------------|----------------|---------------------|----------------------------|---------|------|----------|----------|-------|-----|
| delete | update | upload | create | id | isolate | aliases | country | year | disease | | clonal complex | PorA VR1 | PorA VR2 | FetA VR | ST | PorA VR1 | PorA VR2 | test9 | |
| | | | | 240 | MC58 | Z7176 | UK | 1983 | | Neisseria meningitidis | B | 74 | ST-32 complex/ET-5 complex | 7 | 16-2 | F1-5 | 7 | 16-2 | add |
| | | | | | | | | | | | | update | update | update | | update | update | | |

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

species: *Neisseria meningitidis* update history: 137 updates [show details](#)

Publications (4)

- 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 [79 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](#)
- Stabler RA, Marsden GL, Witney AA, Li Y, Bentley SD, Tang CM, Hinds J (2005). Identification of pathogen-specific genes through microarray analysis of pathogenic and commensal *Neisseria* species. *Microbiology* **151**:2907-22 [3 isolates](#)
- Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Hickey EK, Haft DH, Salzberg SL, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parksey DS, Blair E, Clifton H, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Vamathevan J, Gill J, Scarlato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC (2000). Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**:1809-15 [1 isolate](#)

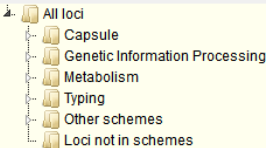
Sequence bin

contigs: 1 detailed breakdown: [Display](#)

length: 2272360 bp

loci tagged: 1283

Schemes and loci


 All loci
 Capsule
 Genetic Information Processing
 Metabolism
 Typing
 Other schemes
 Loci not in schemes

Navigate and select schemes within tree to display allele designations

Click the 'Renumber' button:

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

Logged in: *Keith Jolley (keith)*. [Log out](#) | [Change password](#) Toggle:

Sequence bin for MC58

Contig summary statistics

- Number of contigs: 1
- Length: 2272360
- [Download sequences \(FASTA format\)](#)
- [Download sequences with annotations \(EMBL format\)](#)

| Sequence | Sequencing method | Original designation | Length | Comments | float test | SRA accession | test attribute | test date | test int | Locus | Start | End | Direction | EMBL format | Artemis | Renumber |
|----------|-------------------|----------------------|---------|--------------|------------|---------------|----------------|-----------|----------|----------|-------|-------|-----------|-------------|---------|--------------------------|
| 1 | Sanger | | 2272360 | whole genome | | | | | | NEIS2140 | 502 | 897 | ← | EMBL | Artemis | Renumber |
| | | | | | | | | | | NEIS2141 | 918 | 2312 | ← | | | |
| | | | | | | | | | | NEIS2142 | 2517 | 3161 | ← | | | |
| | | | | | | | | | | NEIS2143 | 3158 | 3511 | ← | | | |
| | | | | | | | | | | NEIS2144 | 3635 | 4117 | → | | | |
| | | | | | | | | | | NEIS2145 | 4311 | 4961 | → | | | |
| | | | | | | | | | | NEIS2146 | 4958 | 5875 | → | | | |
| | | | | | | | | | | NEIS2147 | 5936 | 6214 | → | | | |
| | | | | | | | | | | NEIS2148 | 6281 | 7492 | ← | | | |
| | | | | | | | | | | (pgk) | | | | | | |
| | | | | | | | | | | NEIS2149 | 7573 | 8826 | ← | | | |
| | | | | | | | | | | NEIS2150 | 9346 | 10317 | ← | | | |
| | | | | | | | | | | NEIS2151 | 10350 | 10811 | ← | | | |
| | | | | | | | | | | NEIS2152 | 10840 | 12177 | ← | | | |
| | | | | | | | | | | (kdtA) | | | | | | |
| | | | | | | | | | | NEIS2153 | 12174 | 13622 | ← | | | |
| | | | | | | | | | | NEIS0001 | 15221 | 16144 | ← | | | |

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Renumber locus genome positions based on tagged sequences

You have selected to renumber the genome positions set in the locus table based on the tagged sequences in sequence id#1.

Option: ☐ Remove positions for loci not tagged in this sequence Action: **Renumber**

The following designations will be made:

| Locus | Existing genome position | New genome position |
|----------|--------------------------|---------------------|
| NEIS2140 | 2181973 | 502 |
| NEIS2141 | 2182389 | 918 |
| NEIS2142 | 2183899 | 2517 |
| NEIS2143 | 2184552 | 3158 |
| NEIS2144 | 2185029 | 3635 |
| NEIS2145 | 2185705 | 4311 |
| NEIS2146 | 2186352 | 4958 |
| NEIS2147 | 2187330 | 5936 |
| NEIS2148 | 2187675 | 6281 |
| NEIS2149 | 2188934 | 7573 |
| NEIS2150 | 2190551 | 9346 |
| NEIS2151 | 2191558 | 10350 |
| NEIS2152 | 2192048 | 10840 |
| NEIS2153 | 2193382 | 12174 |
| NEIS0001 | 1261 | 15221 |
| NEIS0210 | 209923 | 17229 |
| piIS | 3271 | 18127 |
| NEIS0004 | 4069 | 23904 |
| NEIS0005 | 4476 | 24311 |
| NEIS0007 | 5843 | 25679 |

5.23 Defining composite fields

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

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


[serogroup]: P1.[PorA_VR1],[PorA_VR2]: [FetA_VR]: ST-[ST] ([clonal_complex])

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

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

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


| | | | |
|----------------------|------|---|--|
| experiment sequences | | ? | Add links associating sequences to experiments. |
| sequence tags | scan | ? | Tag regions of sequences within the sequence bin with locus information. |



Database configuration

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

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

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.


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

Logged in: **Keith Jolley (keith)**. [Log out](#) | [Change password](#)
Toggle: 

Add new composite field

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record


id: strain_designations name of the field as it will appear in the web interface
position after: isolate field present in the isolate table
main display: ☐ true ☒ false Sets whether to display field in isolate query results table (can be overridden by user preference).
curator: Keith Jolley (keith)
timestamp: 2014-07-08

Action

Reset Submit

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.


| | | | | |
|----------------------|------|--|---|--|
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | scan | | ? | Tag regions of sequences within the sequence bin with locus information. |


Database configuration

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

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

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


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


Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#)

Update or delete composite field

1 composite field defined.

| Delete | Update | field name | position after | main display | definition | missing data |
|------------------------|------------------------|--------------------|----------------|--------------|---|----------------------------------|
| Delete | Update | strain_designation | isolate | false | [serogroup]: P1.[PorA_VR1].[PorA_VR2]: [FetA_VR]: ST-[scheme 1:ST] ([scheme 1:clonal_complex]) | 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.


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

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#)

Update composite field - strain designation

position after:
 main display:
 curator: Keith Jolley
 datestamp: 2014-04-09

| field | empty value | regex | curator | datestamp | delete | edit | move |
|-----------------------------|-------------|-----------------------------|--------------|------------|---------------------------------------|-------------------------------------|---|
| serogroup [isolate field] | ND | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| : P1. | | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| PorA_VR1 [locus] | ND | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| : | | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| PorA_VR2 [locus] | ND | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| : | | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| FeiA_VR [locus] | F-ND | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| : ST- | | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| ST [MLST field] | ND | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| (| | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
| clonal_complex [MLST field] | - | s/ST-(\S+) complex.*/cc\$1/ | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |
|) | | | Keith Jolley | 2009-11-12 | <input type="button" value="delete"/> | <input type="button" value="edit"/> | <input type="button" value="up"/> <input type="button" value="down"/> |

Add new field:

curator: Keith Jolley
 datestamp: 2014-07-08

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

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

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

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


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

5.24 Extended provenance attributes (lookup tables)

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

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

| | | | | |
|----------------------|------|--|---|--|
| experiment sequences | | | ? | Add links associating sequences to experiments. |
| sequence tags | scan | | ? | Tag regions of sequences within the sequence bin with locus information. |


Database configuration

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

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

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

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


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

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


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - Neisseria PubMLST

 Add, update or delete records

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

 Database configuration

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

Download the Excel template:

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [2](#)

Batch insert isolate value extended attributes

This page allows you to upload isolate field extended attribute value data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)

Paste in tab-delimited text (include a field header line).

Action

[Back](#)

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

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

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

5.25 Sequence bin attributes


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

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

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

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

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


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)
Toggle: [?](#)

Add new sequence attribute

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

key: read_length
type: integer

curator: Keith Jolley (keith)
timestamp: 2014-07-15
description:

Action

Reset Submit

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

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

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

Batch insert sequences

This page allows you to upload sequence data for a specified isolate record in FASTA format.

If an isolate id is chosen, then all sequences will be associated with that isolate. Alternatively, the isolate id, or any other isolate table field that uniquely defines the isolate, can be named in the identifier rows of the FASTA file. This allows data for multiple isolates to be uploaded.

Please note that you can reach this page for a specific isolate by [querying isolates](#) and then clicking 'Upload' within the isolate table.

Please fill in the following fields - required fields are marked with an exclamation mark (!).

Paste in sequences in FASTA format:

Attributes

isolate id: !

identifier field:

sender: !

method:

run id:

assembly id:

read length:

Options

☐ Don't insert sequences shorter than bps.

Link to experiment:

Alternatively upload FASTA file

Select FASTA file: No file selected.

Action

[Back](#)

5.26 Checking external database configuration settings

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

experiment sequences
sequence tags [scan](#) ? Tag regions of sequences within the sequence bin with locus information.

Database configuration

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

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

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.

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

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

Configuration check - Neisseria PubMLST

Helper applications

| Program | Path | Installed | Executable |
|------------------|--|-----------|------------|
| EMBOSS infoalign | /usr/bin/infoalign | ok | ok |
| EMBOSS sixpack | /usr/bin/sixpack | ok | ok |
| EMBOSS stretcher | /usr/bin/stretcher | ok | ok |
| blastn | /usr/local/ncbi-blast+/bin/blastn | ok | ok |
| blastp | /usr/local/ncbi-blast+/bin/blastp | ok | ok |
| blastx | /usr/local/ncbi-blast+/bin/blastx | ok | ok |
| ipress | /usr/bin/ipress | ok | ok |
| mafft | /usr/local/bin/mafft | ok | ok |
| makeblastdb | /usr/local/ncbi-blast+/bin/makeblastdb | ok | ok |
| mogrify | /usr/bin/mogrify | ok | ok |
| muscle | /usr/bin/muscle | ok | ok |
| tblastx | /usr/local/ncbi-blast+/bin/tblastx | ok | ok |

Locus databases

| Locus | Database | Host | Port | Table | Primary id field | Secondary id field | Secondary id field value | Sequence field | Database accessible | Sequence query | Sequences assigned |
|-------------------|---------------------------------|-----------|------|-----------|------------------|--------------------|--------------------------|----------------|---------------------|----------------|--------------------|
| '16S rDNA | pubmlst_bigsdb_neisseria_seqdef | localhost | 5432 | sequences | allele_id | locus | '16S_rDNA | sequence | ok | ok | 197 |
| 'porA | pubmlst_bigsdb_neisseria_seqdef | localhost | 5432 | sequences | allele_id | locus | 'porA | sequence | ok | ok | 164 |
| 'porB | pubmlst_bigsdb_neisseria_seqdef | localhost | 5432 | sequences | allele_id | locus | 'porB | sequence | ok | ok | 695 |
| 'tpiF | pubmlst_bigsdb_neisseria_seqdef | localhost | 5432 | sequences | allele_id | locus | 'tpiF | sequence | ok | ok | 109 |
| BACT000001 (rpsA) | bigsdb_multispecies_seqdef | localhost | 5432 | sequences | allele_id | locus | BACT000001 | sequence | ok | ok | 8199 |
| BACT000002 (rpsB) | bigsdb_multispecies_seqdef | localhost | 5432 | sequences | allele_id | locus | BACT000002 | sequence | ok | ok | 6567 |
| BACT000003 (rpsC) | bigsdb_multispecies_seqdef | localhost | 5432 | sequences | allele_id | locus | BACT000003 | sequence | ok | ok | 5965 |
| BACT000004 (rpsD) | bigsdb_multispecies_seqdef | localhost | 5432 | sequences | allele_id | locus | BACT000004 | sequence | ok | ok | 6195 |
| BACT000005 (rpsE) | bigsdb_multispecies_seqdef | localhost | 5432 | sequences | allele_id | locus | BACT000005 | sequence | ok | ok | 5707 |
| BACT000006 (rpsF) | bigsdb_multispecies_seqdef | localhost | 5432 | sequences | allele_id | locus | BACT000006 | sequence | ok | ok | 4918 |
| BACT000007 (rpsG) | bigsdb_multispecies_seqdef | localhost | 5432 | sequences | allele_id | locus | BACT000007 | sequence | ok | ok | 5667 |
| BACT000008 (rpsH) | bigsdb_multispecies_seqdef | localhost | 5432 | sequences | allele_id | locus | BACT000008 | sequence | ok | ok | 5260 |

Any problems will be highlighted with a red X.

5.27 Exporting table configurations

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

Experiment sequences

sequence tags scan ? Add links associating sequences to experiments.
 Tag regions of sequences within the sequence bin with locus information.

Database configuration

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

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

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Query schemes for Neisseria PubMLST database

Please enter your search criteria below (or leave blank and submit to return all records).

Search criteria: id = + [?](#) Display: Order by: id ascending Display: 25 records per page [?](#)

Filter query by:

isolate display: [?](#)
 main display: [?](#)
 query field: [?](#)
 query status: [?](#)
 analysis: [?](#)
 allow missing loci: [?](#)
 curator: [?](#)
 scheme: MLST [?](#)

Action:

Click the button 'Export configuration/data'.

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Query schemes for Neisseria PubMLST database

Please enter your search criteria below (or leave blank and submit to return all records).

Search criteria: id = + [?](#) Display: Order by: id ascending Display: 25 records per page [?](#)

Filter query by:

isolate display: [?](#)
 main display: [?](#)
 query field: [?](#)
 query status: [?](#)
 analysis: [?](#)
 allow missing loci: [?](#)
 curator: [?](#)
 scheme: MLST [?](#)

Action:

1 record returned.

Delete Database configuration

| Delete | Update | id | description | dbase name | dbase host | dbase port | dbase table | isolate display* | main display* | query field* | query status* | analysis* | display order | allow missing loci |
|------------------------|------------------------|----|-------------|---------------------------------|------------|------------|-------------|------------------|---------------|--------------|---------------|-----------|---------------|--------------------|
| Delete | Update | 1 | MLST | pubmlst_bigsdb_neisseria_seqdef | | | mv_scheme_1 | true | true | true | true | true | 1 | |

* Default values are displayed for this field. These may be overridden by user preference.

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

```
schemes
-----
id description dbase_name dbase_host dbase_port dbase_user dbase_password dbase_table isolate_display
1 MLST pubmlst_bigsdb_neisseria_seqdef mv_scheme_1 1 1 1 1 1 1 2 2012-03-22

scheme_members
-----
scheme_id locus profile_name field_order curator datestamp
1 abcZ 1 2 2009-11-12
1 adk 2 2 2009-11-12
1 aroE 3 2 2009-11-12
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
-----
scheme_id field type primary_key description field_order url isolate_display main_display qu
1 ST integer 1 1 /cgi-bin/bigsdb/bigsdb.pl?page=profileInfo&db=pubmlst_neisseria_seqdef&scheme
1 clonal_complex text 0 2 1 1 1 1 2 2009-11-16
```

5.28 Authorizing third-party client software to access authenticated resources

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

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

```
create_client_credentials.pl --help

NAME
    create_client_credentials.pl - Generate and populate
    authentication database with third party application (API client)
    credentials.

SYNOPSIS
    create_client_credentials.pl --application NAME [options]

OPTIONS
    -a, --application NAME
        Name of application.

    -d, --deny
        Set default permission to 'deny'. Permissions for access to specific
        database configurations will have to be set. If not included, the default
        permission will allow access to all resources by the client.

    -h, --help
        This help page.
```

```
-i, --insert
    Add credentials to authentication database.  This will fail if a matching
    application version already exists (use --update in this case to overwrite
    existing credentials).

-u, --update
    Update exisitng credentials in the authentication database.

-v, --version VERSION
    Version of application (optional).
```


Curator's guide

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

6.1 Adding new sender details

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

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


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* PubMLST

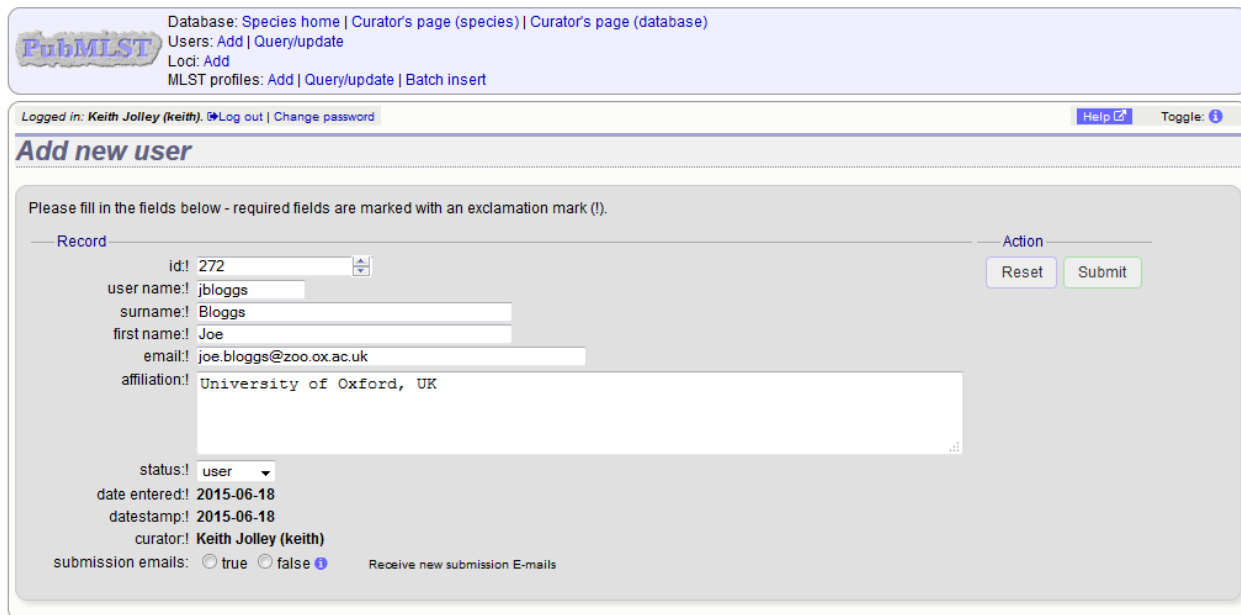

Add, update or delete records

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


Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|---------------|-----|-------------------------------|------------------|--|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |

Enter the user's details in to the form.



Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#) [Toggle](#)

Add new user

Please fill in the fields below - required fields are marked with an exclamation mark (!).

| Record | Action |
|--|--------------|
| id: 272 user name: jbloggs surname: Bloggs first name: Joe email: joe.bloggs@zoo.ox.ac.uk affiliation: University of Oxford, UK status: user date entered: 2015-06-18 datestamp: 2015-06-18 curator: Keith Jolley (Keith) submission emails: <input type="radio"/> true <input checked="" type="radio"/> false ? | Reset Submit |

Receive new submission E-mails


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

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

6.2 Adding new allele sequence definitions


6.2.1 Single allele

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


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* locus/sequence definitions


 Add, update or delete records


| Record type | Add | Batch Add | Update or delete | Comments |
|---|-----|-----------|--|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rpIF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

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

The status can either be:

- Sanger trace checked
- WGS: manual extract (BIGSdb)
- WGS: automated extract (BIGSdb)
- WGS: visually checked
- WGS: automatically checked
- unchecked


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)
[Help](#)
Toggle: 

Add new allele sequence

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

locus:!

allele id:!

sequence:!
TTTGATACCGTTGCGGAAGGTTTGGGCGAAATTCGCGATTATTGCGCCGTTATCATGTGAGCCATGATTGGAGCACAATTCAGACGACGTTTTATTAAAAGAACTCAACGAATTACAACCTCGAAATCGAAGCGAAGGACGGCTGGAAGCTGGATGCGGCAGTCAAGCAGACTTTGGGCGAACTCGGTTTGGCGAAAACGAAAAATCGGCAACCTTTCCGGCGGT CAGAAAAAGCGTGTGCGCTTGGCGCAGGCTTGGGTGCAGAAGCCCGACGTATTGCTGCTGGACGAACCGACCAACCA TTTGGATATTGACGCGATTATCTGGTTGGAAAACCTGCTCAAGGCGTTTGAAGGCAGCTTGGTCGTGATTACCCACG ATCGCCGTTTTTTGGACAATATCGCCACGCGGATTGTGGAACTCGATC

status:!

sender:!

curator:! **Keith Jolley (keith)**

date entered:! **2015-06-18**

datestamp:! **2015-06-18**

comments:

Flags:

 Use Ctrl click to select/deselect multiple choices

PubMed ids:

ENA ids:

Genbank ids:

☐ Override sequence similarity check
☐ Override sequence length check

Action

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

See also:

allele sequence flags


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

6.2.2 Batch adding multiple alleles

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


Upload using a spreadsheet

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


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)


Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)


Database curator's interface - *Neisseria* locus/sequence definitions

 Add, update or delete records

| Record type | Add | Batch Add | Update or delete | Comments |
|---|-----|---------------------------------|--|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rplF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

Download a template Excel file from the following page.


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loc: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)
Toggle: 

Batch insert sequences

This page allows you to upload allele sequence data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- If the locus uses integer allele ids you can leave the allele_id field blank and the next available number will be used.
- The status defines how the sequence was curated. Allowed values are: 'Sanger trace checked', 'WGS: manual extract', 'WGS: automated extract', 'unchecked'
- Sequence flags can be added as a semi-colon (;) separated list
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)
- Please note, some loci have extended attributes which may be required. For affected loci please use the batch insert page specific to that locus:

Reload page specific for locus:

Please select the sender from the list below:

Select sender ... Value will be overridden if you include a sender field in your pasted data.

☒ 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


Paste in tab-delimited text (include a field header line).


Action

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

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

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


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)
Toggle: 

Batch insert sequences

This page allows you to upload allele sequence data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- If the locus uses integer allele ids you can leave the allele_id field blank and the next available number will be used.
- The status defines how the sequence was curated. Allowed values are: 'Sanger trace checked', 'WGS: manual extract', 'WGS: automated extract', 'unchecked'
- Sequence flags can be added as a semi-colon (;) separated list
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)
- Please note, some loci have extended attributes which may be required. For affected loci please use the batch insert page specific to that locus:

Reload page specific for locus: Select ... Reload

Please select the sender from the list below:

Jolley, Keith (keith) Value will be overridden if you include a sender field in your pasted data.

☒ 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

Paste in tab-delimited text (include a field header line).

| locus | allele_id | sequence | status | comments | flags |
|-------|-----------|---|----------------------|----------|-------|
| abcZ | | TTTGATACIGTGGCCGAAGCTTTGGGCGAAATTCGCGATTTATTGCAACGTTATCATCATGTCAGCCATGAGTTGGAAAAAT | | | |
| | | GGITCGAGTGGAGGCTTATGAAAGAGCTCAACGAATTCGCAACTTGAGATCGAAGCGAAGGACGCGCTGGAAAGTTGGATCGC | | | |
| | | GCGGTGAAGCGAGACTTTGGGCGAACTCGGTTTGGCGGAAACGAAAAATCGGCAACCTCTCCGCGGTTCAGAAAAAGCGC | | | |
| | | GTGCGCTTGGCCGAGGCTTTGGGTCGAGAGCCGAGCTTTTCTGCTCGATGAAACGACCAACATTTGGACATCGACGCG | | | |
| | | ATTATTTGGTTGGAAAACTTCTCAAAAGCGTTTGAAGGCGAGCTGGTTTGATTTACCCACGACCGCGTTTTTTGGACAAAT | | | |
| | | ATCGCCACGCGGATTGTGGAACCTCGATC | Sanger trace checked | | |
| abcZ | | TTTGATACIGTGGCCGAAGCTTTGGGCGAAATTCGCGATTTATTGCAACGTTATCATCATGTCAGCCATGAGTTGGAAAAAT | | | |
| | | GGITCGAGTGGAGGCTTATGAAAGAGCTCAACGAATTCGCAACTTGAGATCGAAGCGAAGGACGCGCTGGAAAGTTGGATCGC | | | |
| | | GCGGTGAAGCGAGACTTTGGGCGAACTCGGTTTGGCGGAAACGAAAAATCGGCAACCTCTCCGCGGTTCAGAAAAAGCGC | | | |
| | | GTGCGCTTGGCCGAGGCTTTGGGTCGAGAGCCGAGCTTTTCTGCTCGATGAAACGACCAACATTTGGACATCGACGCG | | | |
| | | ATTATTTGGTTGGAAAACTTCTCAAAAGCGTTTGAAGGCGAGCTGGTTTGATTTACCCACGACCGCGTTTTTTGGACAAAT | | | |
| | | ATCGCCACGCGGATTGTGGAACCTCGATC | Sanger trace checked | | |

[Back](#)

Action
Reset Submit

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: ?

Batch insert sequences

Import status
 Sender: Keith Jolley
 No obvious problems identified so far.

[Import data](#)

Data to be imported
 The following table shows your data. Any field with red text has a problem and needs to be checked. *Note: valid sequence flags are displayed with a red background not red text.*

| locus | allele_id | sequence | status | sender | curator | date_entered | datestamp | comments | flags |
|-------|-----------|--|----------------------|--------|---------|--------------|------------|----------|-------|
| abcZ | 688 | TTTGATACTGTTGCCGAAGG ... GCGGATTGTGCGAACTCGATC | Sanger trace checked | 2 | 2 | 2014-07-09 | 2014-07-09 | | |
| abcZ | 689 | TTTGATACTGTTGCCGAAGC ... GCGGATTGTGCGAACTCGATC | Sanger trace checked | 2 | 2 | 2014-07-09 | 2014-07-09 | | |

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: ?

Batch insert sequences

Import status

| Primary key | Problem(s) |
|-------------------------|--|
| locus: abcZ; allele_id: | Sequence contains non nucleotide (G A T C) 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 are displayed with a red background not red text.*

| locus | allele_id | sequence | status | sender | curator | date_entered | datestamp | comments | flags |
|-------|-----------|--|----------------------|--------|---------|--------------|------------|----------|-------|
| abcZ | 688 | TTTGATACTGTTGCCGAAGG ... GCGGATTGTGCGAACTCGATC | Sanger trace checked | 2 | 2 | 2014-07-09 | 2014-07-09 | | |
| abcZ | 689 | TTTGATACTGTTGCCGAAGC ... GCGGATTGTGCGAACTCGATC | Sanger trace checked | 2 | 2 | 2014-07-09 | 2014-07-09 | | |

Upload using a FASTA file


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

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)


Database curator's interface - *Neisseria* locus/sequence definitions

 Add, update or delete records

| Record type | Add | Batch Add | Update or delete | Comments |
|---|-----|-----------------|--|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rpIF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: 

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:

status:

sender:

sequence (FASTA):

```
>isolate1
TTTGATACTGTTGCCGAAGGTTTGGGCGAAATTCGCGATTATTGCAACGTTATCATCAT
GTCAGCCATGAGTTGGAAAATGGTTCGAGTGAGGCCCTTATTGAAAGAGCTCAACGAATTG
CAACTTGAGATCGAAGCGAAGGACGGCTGSAAGTTGGATGCGGCGGTGAAGCAGACTTTG
GGCGAACTCGGTTTGGCGAAAACGAAAAATCGGCAACCTCTCCGGCGGTGAGAAAAAG
CGCGTCGCGCTTGGCGCAGGCTTGGGTGCAGAACCCGACGATTGCTGCTCGATGAACCG
ACCAACCATTTGGACATCGACGCGATTATTGGTTGGAAAACCTGCTCAAAGCGTTTGAA
GGCAGCCTGGTTGTGATTACCCACGACCGCGCTTTTGGACATATCGCCACGCGGATT
GTCGAATCTGATC
>isolate2
TTTGATACTGTTGCCGAAGCTTTGGGCGAAATTCGCGATTATTGCAACGTTATCATCAT
```

☐ 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)

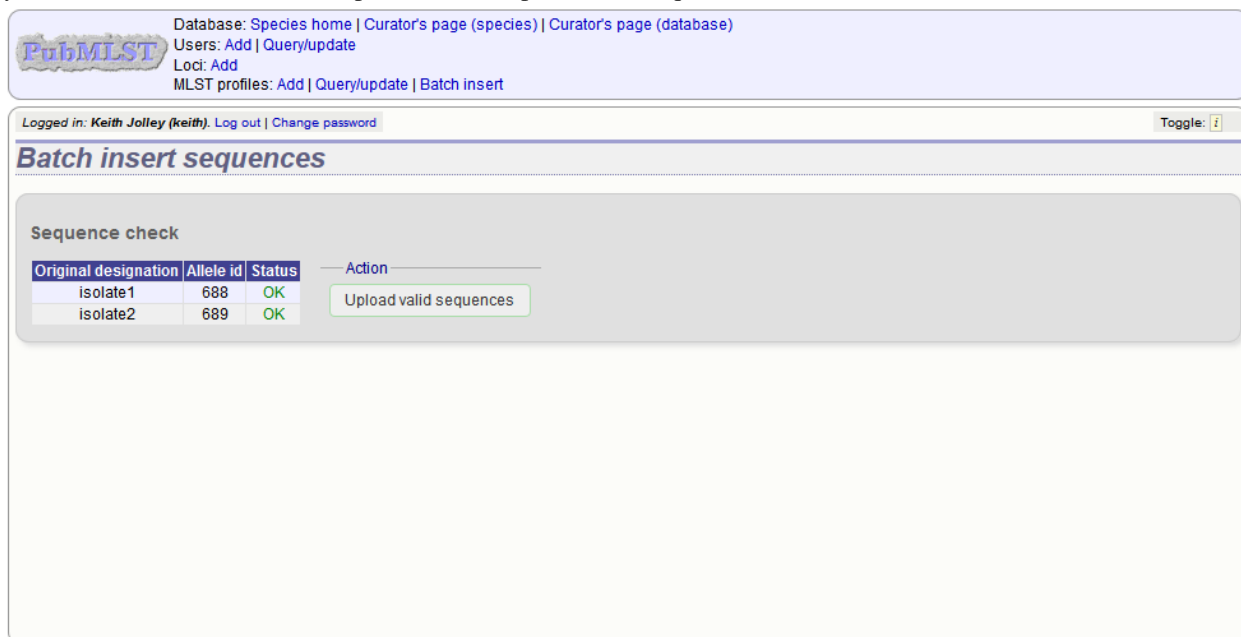
Action

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

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

Click 'Check'.

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



Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: ?

Batch insert sequences

Sequence check

| Original designation | Allele id | Status | Action |
|----------------------|-----------|--------|--|
| isolate1 | 688 | OK | Upload valid sequences |
| isolate2 | 689 | OK | |

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



Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: ?

Batch insert sequences

Sequence check

| Original designation | Allele id | Status | Action |
|----------------------|-----------|--|--|
| isolate1 | 688 | OK | Upload valid sequences |
| isolate2 | 689 | Sequence contains non nucleotide (G A T C) characters. | |

6.3 Updating and deleting allele sequence definitions

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

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

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* locus/sequence definitions

Add, update or delete records

| Record type | Add | Batch Add | Update or delete | Comments |
|---|-----|------------|--|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rpIF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Query sequences for *Neisseria* locus/sequence definitions database

Some loci have additional fields which are not searchable from this general page. Search for these at the [locus-specific query](#) page. Use this page also for access to the sequence analysis or export plugins.

Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a locus that uses integer allele ids using the drop-down list.

Please enter your search criteria below (or leave blank and submit to return all records).

Search criteria: allele id = 4 [+](#) [?](#)

Display: Order by: locus ascending Display: 25 records per page [?](#)

Filter query by: locus: abcZ [?](#) status: [?](#) sender: [?](#) curator: [?](#) allele flag: [?](#)

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.

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Query sequences for *Neisseria* locus/sequence definitions database

Some loci have additional fields which are not searchable from this general page. Search for these at the [locus-specific query](#) page. Use this page also for access to the sequence analysis or export plugins.

Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a locus that uses integer allele ids using the drop-down list.

Please enter your search criteria below (or leave blank and submit to return all records).

Search criteria **Display**

allele id = 4 + [?](#) Order by: locus ascending Display: 25 records per page [?](#)

Filter query by **Action**

locus: abcZ [?](#) Reset Submit

status: [?](#)

sender: [?](#)

curator: [?](#)

allele flag: [?](#)

1 record returned.

[Delete](#) [Database configuration](#) [Flags](#)

[Delete ALL](#) [Export configuration/data](#) [Batch set](#)

| Delete | Update | locus | allele id | sequence | sequence length | status | sender | curator | date entered | timestamp | comments | flags |
|------------------------|------------------------|-----------------------|---------------------------|---|---------------------------------|------------------------|------------------------|-------------------------|------------------------------|---------------------------|--------------------------|-----------------------|
| Delete | Update | abcZ | 4 | TTTGATACCG TTGCGAAGG TTGGGGGAA ATTCTGATT TATTGGCGG TTAATCATCAT GTACGCCATG AGTTGGAAAA TGGTTCGAGT GAGGCTTTGT TGAAAGAACT CAACGAATTG CAACTTGAAA | 433 | Sanger trace checked | Keith Jolley | Man-Suen Chan | 2001-02-07 | 2009-11-11 | | |

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Delete allele sequence

You have chosen to delete the following record:

locus: abcZ

allele id: 4

sequence: TTTGATACCG TTGCGAAGG TTGGGGGAA ATTCTGATT TATTGGCGG TTAATCATCAT GTACGCCATG AGTTGGAAAA TGGTTCGAGT GAGGCTTTGT TGAAAGAACT CAACGAATTG CAACTTGAAA
 TCGAAGCGAA GGACGGCTGG AAACCTGGATG CGGCAGTCAA GCAGACTTIG GGGGAATCG GTTTCGCGGA AAATGAAAAA ATCGGCAACC TTTCGCGCGG TCAGAAAAAG CGCGTCGCTT TGGCTCAGGC
 TTGGGTGCAA AAGCCCGACG TATTGCTGCT GGACGAGCCG ACCAACCATT TGGATATCGA CGGATTATT TGGCTGGAAA ATCTGCTCAA AGCGTTTGAA GGCAGCTTGG TTGTGATTAC CCACGACCGC
 CGTTTTTTGG ACAATATCGC CACGCGGATT GTCGAACTCG ATC

status: Sanger trace checked

sender: Keith Jolley

curator: Man-Suen Chan

date entered: 2001-02-07


timestamp: 2009-11-11


comments:

[Delete!](#)

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

Note: Adding flags and comments to an allele record requires that this feature is enabled in the [database configuration](#).


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: **Keith Jolley (keith)**. [Log out](#) | [Change password](#)
Toggle: 

Update allele sequence

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

locus: **abcZ**
allele id: **4**
sequence:

TTTGATACCG TTGCCGAAGG TTTGGGCGAA
AATCGTGATT TATTGCGCCG TTATCATCAT
GTCAGCCATG AGTTGGAAAA TGGTTCGAGT
GAGGCTTTGT TGAAAGAACT CAACGAATTG
CAACTTTGAA TCGAAGCGAA GGACGCTTGG
AAACTGGATG CGGCAGTCAA GCAGACTTTG
GGGGAACTCG GTTTGCCGGA AAATGAAAA

status: **Sanger trace checked**
sender: **Jolley, Keith (keith)**
curator: **Keith Jolley (keith)**
date entered: **2001-02-07**
datestamp: **2014-07-09**
comments:
Flags:

atypical
contains IS element
downstream fusion
frameshift
internal stop codon

PubMed ids:
ENA ids:
Genbank ids:

Action

Reset Submit

6.4 Updating locus descriptions

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


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

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)


Database curator's interface - *Neisseria* locus/sequence definitions

 Add, update or delete records

| Record type | Add | Batch Add | Update or delete | Comments |
|---|-----|------------|--|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rpIF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: 

Query locus descriptions for *Neisseria* locus/sequence definitions database

Please enter your search criteria below (or leave blank and submit to return all records).

Search criteria: locus = NEIS0620 + i

Display: Order by: locus ascending
 Display: 25 records per page i

Filter query by: Action

Reset Submit

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#) Toggle: [i](#)

Query locus descriptions for *Neisseria* locus/sequence definitions database

Please enter your search criteria below (or leave blank and submit to return all records).

Search criteria **Display**

locus: = Order by: locus ascending

☒ **Filter query by** Display: 25 records per page

locus: **Action**

curator: Reset Submit

common name:

Click 'Submit'.

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#) Toggle: [i](#)

Query locus descriptions for *Neisseria* locus/sequence definitions database

Please enter your search criteria below (or leave blank and submit to return all records).

Search criteria **Display**

locus: = Order by: locus ascending

☒ **Filter query by** Display: 25 records per page

locus: **Action**

curator: Reset Submit

common name:

1 record returned.

[Delete](#) [Database configuration](#)

[Delete ALL](#) [Export configuration/data](#)

| Delete | Update | locus | full name | product | description | curator | timestamp |
|------------------------|------------------------|----------|-----------|-------------------------------------|---|--------------|------------|
| Delete | Update | 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:

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Update locus description

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record **Action**

locus: **NEIS0620** [Reset](#) [Submit](#)

curator: **Keith Jolley (keith)**

timestamp: **2014-11-04**

full name:

product: **malate oxidoreductase (EC 1.1.1.38)**

description: **Final step in TCA cycle producing oxaloacetate.**

aliases: **NMA0870**
NMB0671
NMC0620

PubMed ids: **14917678**

links: **http://www.enzyme-database.org**
/query.php?ec=1.1.1.38|EC 1.1.1.38

(Format: URL|description)

- 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.

6.5 Adding new scheme profile definitions


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

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


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* locus/sequence definitions

 Add, update or delete records

| Record type | Add | Batch Add | Update or delete | Comments |
|---|---|------------|--|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rplF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

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.

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle:

Add new MLST profile

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

ST: ! 10880
 abcZ: ! 2
 adk: ! 3
 aroE: ! 4
 fumC: ! 122
 gdh: ! 8
 pdhC: ! 4
 pgm: ! 6
 sender: !
 clonal_complex:
 curator: ! **Keith Jolley (keith)**
 date_entered: ! 2014-07-10
 datestamp: ! 2014-07-10
 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:

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* locus/sequence definitions

Add, update or delete records

| Record type | Add | Batch Add | Update or delete | Comments |
|---|-----|------------|--|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rpIF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

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

 Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: 


Batch insert MLST profiles


This page allows you to upload profiles 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.
- 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, then you must also provide it for each profile record.
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)

Please paste in tab-delimited text (include a field header line)

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


Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
Users: [Add](#) | [Query/update](#)
Loci: [Add](#)
MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)
Toggle: 

Batch insert MLST profiles

This page allows you to upload profiles 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.
- 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, then you must also provide it for each profile record.
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)

Please paste in tab-delimited text (include a field header line)

| abcZ | adk | aroE | fumC | gdh | pdhC | pgm | clonal_complex |
|------|-----|------|------|-----|------|-----|----------------|
| 2 | 3 | 4 | 122 | 8 | 4 | 6 | |

Parameters
Sender: Jolley, Keith (keith)
Value will be overridden if you include a sender field in your pasted data.
☐ Ignore duplicate profiles

Action
Reset Submit

[Back](#)

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle:

Batch insert MLST profiles

Import status
 Sender: Keith Jolley
 No obvious problems identified so far.

[Import data](#)

Data to be imported
 The following table shows your data. Any field coloured red has a problem and needs to be checked.

| ST | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | clonal_complex | sender | curator | date_entered | timestamp |
|-------|------|-----|------|------|-----|------|-----|----------------|--------|---------|--------------|------------|
| 10880 | 2 | 3 | 4 | 122 | 8 | 4 | 6 | | 2 | 2 | 2014-07-10 | 2014-07-10 |

6.6 Updating and deleting scheme profile definitions

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* locus/sequence definitions

Add, update or delete records

| Record type | Add | Batch Add | Update or delete | Comments |
|---|-----|------------|--|---|
| users | + | ++ | ? | |
| user groups | + | ++ | ? | Users can be members of these groups - use for setting access permissions. |
| user group members | + | ++ | ? | Add users to groups for setting access permissions. |
| curator permissions | | | ? | Set curator permissions for individual users - these are only active for users with a status of 'curator' in the users table. |
| locus descriptions | + | ++ | ? | |
| scheme curator control list | + | ++ | query batch | Define which curators can add or update profiles for particular schemes. |
| locus curator control list | + | ++ | query batch | Define which curators can add or update sequences for particular loci. |
| sequences (all loci) | + | ++ FASTA | ? | |
| accession number links | + | ++ | ? | Associate sequences with Genbank/EMBL accession number. |
| PubMed links (to sequences) | + | ++ | ? | |
| Bexsero Antigen Sequence Typing (BAST) profiles | + | ++ | query browse list batch update | |
| MLST profiles | + | ++ | query browse list batch update | |
| rpIF species profiles | + | ++ | query browse list batch update | |
| PubMed links (to profiles) | + | ++ | | |

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Query/update profiles - *Neisseria* locus/sequence definitions

Schemes
 Please select the scheme you would like to query:

MLST

Locus/scheme fields **Display/sort options**

ST Order by: ST
 Display: 25 records per page

Filter query by **Action**

clonal complex:

1 record returned.

[Delete](#)

| Delete | Update | ST | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | clonal complex |
|------------------------|------------------------|------|------|-----|------|------|-----|------|-----|----------------|
| Delete | Update | 4563 | 2 | 7 | 6 | 13 | 9 | 18 | 8 | ST-167 complex |

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

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Delete profile

You have chosen to delete the following record:

```

scheme id: 1) MLST
  ST: 4563
  abcZ : 2
  adk : 7
  aroE : 6
  fumC : 13
  gdh : 9
  pdhC : 18
  pgm : 8
clonal_complex : ST-167 complex
  sender: Ana-Belen Ibarz-Pavon
  curator: Keith Jolley
  date entered: 2005-03-03
  datestamp: 2009-11-11
  
```

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'.

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#) Toggle:

Update profile

Record **Action**

Update your record as required - required fields are marked with an exclamation mark (!): [Reset](#) [Submit](#)

ST: ! 4563
 abcZ: ! 2
 adk: ! 7
 aroE: ! 6
 fumC: ! 13
 gdh: ! 9
 pdhC: ! 18
 pgm: ! 8

clonal_complex: ST-167 complex
 sender: ! Ibarz-Pavon, Ana-Belen (aibarz)
 curator: ! Keith Jolley (keith)
 date_entered: ! 2005-03-03
 datestamp: ! 2014-07-10
 PubMed ids:

6.7 Adding isolate records

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

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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* PubMLST

Add, update or delete records

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

Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-------------------|-----|-------------------------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for taqoano to specific repetitive loci. |

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.

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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#) Toggle:

Add new isolate

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Isolate fields

id: 30457

isolate: J323_2

country: UK

species: *Neisseria meningitidis*

sender: Jolley, Keith (keith)

curator: **Keith Jolley (keith)**

date entered: 2014-07-10

date stamp: 2014-07-10

region:

year: 2014

epidemiological year:

age yr:

age mth:

sex:

disease: meningitis

source: CSF

epidemiology:

serogroup: B

MLEE designation:

serotype:

sero subtype:

ET no:

penicillin:

penicillin range:

amoxicillin:

sulphonamide:

ceftriaxone:

Allele designations

MLST

| | | | |
|------|------|------|------|
| abcZ | adk | aroE | fumC |
| gdh | pdhC | pgm | |

Finotyping antigens

| | | |
|----------|----------|---------|
| PorA VR1 | PorA VR2 | FetA VR |
| | | |

Action

[Reset](#) [Submit](#)

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.

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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* PubMLST

Add, update or delete records

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

Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-------------------|-----|-----------|------------------|--|
| loci | + | ++ | ? | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | ? | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | ? | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for taqoia to specific repetitive loci. |

Download a submission template in Excel format from the link.

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle:

Batch insert isolates

This page allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.
- Enter references for your isolates as a semi-colon (;) separated list of PubMed ids (non integer ids will be ignored).
- You can also upload allele fields along with the other isolate data - simply create a new column with the locus name. These will be added with a confirmed status and method set as 'manual'.
- You can choose whether or not to include an id number field - if it is omitted, the next available id will be used automatically.
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)**

Please select the sender from the list below:


Select sender ... Value will be overridden if you include a sender field in your pasted data.


Paste in tab-delimited text (include a field header line).

Action
 Reset Submit

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

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


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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)
Toggle: 

Batch insert isolates

This page allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.
- Enter references for your isolates as a semi-colon (;) separated list of PubMed ids (non integer ids will be ignored).
- You can also upload allele fields along with the other isolate data - simply create a new column with the locus name. These will be added with a confirmed status and method set as 'manual'.
- You can choose whether or not to include an id number field - if it is omitted, the next available id will be used automatically.

- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)

Please select the sender from the list below:


Jolley, Keith (keith) Value will be overridden if you include a sender field in your pasted data.


Paste in tab-delimited text (include a field header line).

isolate aliases references country region year epidemiological_year
age_yr age_mth sex disease source epidemiology species serogroup
MLEE_designation serotype sero_subtype ET_no penicillin
penicillin_range amoxicillin sulphonamide ceftriaxone
ceftriaxone_range chloramphenicol chloramphenicol_range cefotaxime
cefotaxime_range rifampicin rifampicin_range ciprofloxacin
ciprofloxacin_range pending_assembly assembly_status ENA_accession
private_project comments abcZ adk aroF fumC qdh pdhC
pgm
J392_1 UK 2014 12 male
meningitis and septicaemia CSF Neisseria meningitidis
B

Action
Reset Submit

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


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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)
Toggle: 

Batch insert isolates

Import status
Sender: Keith Jolley
No obvious problems identified so far.

Data to be imported
The following table shows your data. Any field with red text has a problem and needs to be checked.

| id | isolate | aliases | references | country | region | year | epidemiological_year | age_yr | age_mth | sex | disease | source | epidemiology | species | serogroup | MLEE_desig |
|-------|---------|---------|------------|---------|--------|------|----------------------|--------|---------|------|----------------------------|--------|--------------|------------------------|-----------|------------|
| 30457 | J392_1 | | | UK | | 2014 | | 12 | | male | meningitis and septicaemia | CSF | | Neisseria meningitidis | B | |

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [Z](#)

Batch insert isolates

Import status

| Primary key | Problem(s) |
|-------------|---|
| id: 30457 | species 'Neisseria meningitidis' is not on the list of allowed values for this field. |

Data to be imported

The following table shows your data. Any field with red text has a problem and needs to be checked.

| id | isolate | aliases | references | country | region | year | epidemiological_year | age_yr | age_mth | sex | disease | source | epidemiology | species | serogroup | MLEE_desig |
|-------|---------|---------|------------|---------|--------|------|----------------------|--------|---------|------|----------------------------|--------|--------------|------------------------|-----------|------------|
| 30457 | J392_1 | | | UK | | 2014 | | 12 | | male | meningitis and septicaemia | CSF | | Neisseria meningitidis | B | |

6.8 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.


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - Neisseria PubMLST

 Add, update or delete records

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

 Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-------------------|-----|---------------|------------------|---|
| loci | + | ++ | ? | |
| | | database scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for taocino to specific repetitive loci. |

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.

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

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#)

Toggle: [Field help](#): id [Go](#)

Isolate query/update

Isolate provenance/phenotype fields

Combine with: **AND**

country = USA + [i](#)

year = 2009

Display/sort options

Order by: id ascending

Display: 25 records per page [i](#)

Action

[Reset](#) [Submit](#)

19 records returned. Click the hyperlinks for detailed information.

[Delete](#) [Tag scanning](#) [Projects](#)

[Delete ALL](#) [Scan](#) [Select project...](#) [Link](#)

| Delete | Update | Sequence bin | New version | Isolate fields i | | | | | | | | Seqbin size (bp) | MLST | | Finely anti | |
|------------------------|------------------------|------------------------|------------------------|----------------------------------|---------|---------|---------|------|------------|------------------------|-----------|------------------|------------|----------------------------|------------------------|---------------------|
| | | | | id | isolate | aliases | country | year | disease | species | serogroup | | ST | clonal complex | PorA VR1 | PorA VR2 |
| delete | update | upload | create | 12674 | M18700 | | USA | 2009 | | Neisseria meningitidis | B | 0 | 7301 | ST-32 complex/ET-5 complex | add | ac |
| delete | update | upload | create | 12675 | M18701 | | USA | 2009 | | Neisseria meningitidis | B | 0 | 7302 | | add | ac |
| delete | update | upload | create | 12676 | M18725 | | USA | 2009 | | Neisseria meningitidis | B | 0 | 7303 | | add | ac |
| delete | update | upload | create | 13090 | M19024 | PA09015 | USA | 2009 | meningitis | Neisseria meningitidis | B | 0 | 7575 | | add | ac |
| delete | update | upload | create | 14627 | M20918 | | USA | 2009 | | Neisseria meningitidis | A | 1718070 | 4789, 7980 | ST-5 complex/subgroup III | update | upd |
| delete | update | upload | create | 14998 | M21319 | | USA | 2009 | invasive | Neisseria | B | 0 | 8250 | ST-60 complex | add | ac |

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Delete isolate

You have selected to delete the following record:

Provenance/meta data

| | | |
|--|---|---|
| id: 14627 | source: CSF | update history: 26 updates show details |
| isolate: M20918 | species: <i>Neisseria meningitidis</i> | date entered: 2010-01-26 |
| strain designation: A: P1.20.9: F3-1: ST- 4789.7980 (cc5) | serogroup: A | datestamp: 2014-06-17 |
| country: USA | sender: Xin Zhao, Novartis (formerly at US CDC) | |
| continent: North America | curator: Carina Brehony, University of Oxford, UK (E-mail: carina.brehony@zoo.ox.ac.uk) | |
| region: IA | | |
| year: 2009 | | |

Publication (1)

- Kislyuk AO, Katz LS, Agrawal S, Hagen MS, Conley AB, Jayaraman P, Nelakuditi V, Humphrey JC, Sammons SA, Govil D, Mair RD, Tatti KM, Tondella ML, Harcourt BH, Mayer LW, Jordan IK (2010). A computational genomics pipeline for prokaryotic sequencing projects. *Bioinformatics* **26**:1819-26 [18 isolates](#)

Sequence bin

| | |
|--------------------------|---|
| contigs: 2049 | N90: 456 |
| total length: 1718070 bp | N95: 324 |
| max length: 5436 bp | loci tagged: 979 |
| mean length: 839 bp | detailed breakdown: Display |
| N50: 1155 | |

Action

[Delete](#)

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

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#)

Update isolate

Isolate fields

id: 14627 ⓘ
 isolate: M20918 ⓘ
 country: USA ⓘ
 species: *Neisseria meningitidis* ⓘ
 sender: Zhao, Xin (xinzha) ⓘ
 curator: [Keith Jolley \(keith\)](#) ⓘ
 date entered: 2010-01-26 ⓘ
 datestamp: 2015-07-02 ⓘ
 region: IA ⓘ
 year: 2009 ⓘ
 epidemiological year: ⓘ ⓘ
 age yr: ⓘ ⓘ
 age mth: ⓘ ⓘ
 sex: ⓘ
 disease: ⓘ
 source: CSF ⓘ
 epidemiology: ⓘ
 serogroup: A ⓘ
 MLEE designation: ⓘ
 serotype: ⓘ
 sero subtype: ⓘ
 ET no: ⓘ
 penicillin: ⓘ
 penicillin range: ⓘ
 amoxicillin: ⓘ

Loci

All loci

- Genetic Information Processing
- Genomic islands
- Lineage Schemes
- Metabolism
- Pilin
- Typing
- Other schemes
- Loci not in schemes

Navigate and select schemes within tree to display allele designations
 Locus: '16S_rDNA' [Add/update](#)

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

next to a displayed locus.

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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)

[Help](#)

Update isolate

Isolate fields

id: **14627**
isolate: M20918
country: USA
species: *Neisseria meningitidis*
sender: Zhao, Xin (xinzhao)
curator: **Keith Jolley (keith)**
date entered: 2010-01-26
datestamp: 2015-07-02
region: IA
year: 2009
epidemiological year:
age yr:
age mth:
sex:
disease:
source: CSF
epidemiology:
serogroup: A
MLEE designation:
serotype:
sero subtype:
ET no:
penicillin:
penicillin range:
amoxicillin:
amoxicillin range:

Loci

All loci
Genetic Information Processing
Genomic islands
Lineage Schemes
Metabolism
Pilin
Typing
Other schemes
Loci not in schemes

Navigate and select schemes within tree to display allele designations
Locus: '16S_rDNA' [Add/update](#)

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

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)

[Help](#)

Isolate fields

id: **14627**
isolate: M20918
country: USA
species: *Neisseria meningitidis*
sender: Zhao, Xin (xinzhao)
curator: **Keith Jolley (keith)**
date entered: 2010-01-26
datestamp: 2015-07-02
region: IA
year: 2009
epidemiological year:
age yr:
age mth:
sex:
disease:
source: CSF
epidemiology:
serogroup: A
MLEE designation:
serotype:
sero subtype:
ET no:
penicillin:
penicillin range:
amoxicillin:
amoxicillin range:
sulphonamide:

Loci

All loci
Genetic Information Processing
Genomic islands
Lineage Schemes
Metabolism
Pilin
Typing
MLST
Finetyping antigens
Antigen genes
eMLST (20 locus partial genes)
eMLST (20 locus whole genes)
N. meningitidis cgMLST v1.0
Ribosomal MLST
rplF species
Other schemes
Loci not in schemes

MLST

| | | | | | | |
|------------|------------|------------|------------|------------|-----------------|-------------|
| abcZ | adk | aroE | fumC | gdh | pdhC | pgm |
| 1 S update | 1 S update | 2 S update | 1 S update | 3 S update | 6, 334 S update | 19 S update |

ST
4789, 7980
clonal complex
ST-5 complex/subgroup III

Locus: '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'.

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#)

Update isolate

Isolate fields

id: 14627

isolate: M20918

country: USA

species: *Neisseria meningitidis*

sender: Zhao, Xin (xinzhao)

curator: Keith Jolley (keith)

date entered: 2010-01-26

datestamp: 2015-07-02

region: IA

year: 2009

epidemiological year:

age yr:

age mth:

sex:

disease:

source: CSF

epidemiology:

serogroup: A

MLEE designation:

serotype:

sero subtype:

ET no:

penicillin:

penicillin range:

amoxicillin:

sulphonamide:

Loci

All loci

- Genetic Information Processing
- Genomic islands
- Lineage Schemes
- Metabolism
- Pilin
- Typing
- Other schemes
- Loci not in schemes

Navigate and select schemes within tree to display allele designations

Locus: abcZ [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.

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Update abcZ allele for isolate 14627

Provenance/meta data

id: 14627

isolate: M20918

strain designation: A: P1.20.9: F3-1: ST-4789.7980 (cc5)

country: USA

region: IA

year: 2009

source: CSF

species: *Neisseria meningitidis*

serogroup: A

sender: Xin Zhao

curator: Carina Brehony

update history: [26 updates](#) [show details](#)

date entered: 2010-01-26

datestamp: 2014-06-17

Locus: abcZ

[Add new allele designation](#)

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

isolate id: 14627

locus: abcZ

allele id: 5

sender: Jolley, Keith (keith)

status: confirmed

method: manual

curator: Keith Jolley (keith)

datestamp: 2014-07-10

date entered: 2014-07-10

comments:

Action

[Reset](#) [Submit](#)

Existing designations

| Update | Delete | allele id | sender | status | method | comments |
|------------------------|------------------------|-----------|----------|-----------|--------|----------|
| Update | Delete | 1 | Xin Zhao | confirmed | manual | |

Update other loci:

Locus: abcZ [Add/update](#)

6.9 Batch updating multiple isolate records

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

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

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

Database curator's interface - Neisseria PubMLST

 Add, update or delete records

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

 Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-----------------------------------|-----|-----------|------------------|---|
| loci | + | ++ | ? | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attributes | + | ++ | ? | Define additional attributes to associate with values of a particular isolate record field. |

Prepare your update data in 3 columns in a spreadsheet:

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

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

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

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

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

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

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

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#)

Batch isolate update

This page allows you to batch update provenance fields or allele designations for multiple isolates.

- The first line, containing column headings, will be ignored.
- The first column should be the isolate id (or unique field that you are selecting isolates on). If a secondary selection field is used (so that together the combination of primary and secondary fields are unique), this should be entered in the second column.
- The next column should contain the field/focus name and then the final column should contain the value to be entered, e.g.

```
id      field  value
2       country USA
2       abcZ   5
```

- The columns should be separated by tabs. Any other columns will be ignored.
- If you wish to blank a field, enter '<blank>' as the value.
- The script is compatible with STARS output files.

Please enter the field(s) that you are selecting isolates on. Values used must be unique within this field or combination of fields, i.e. only one isolate has the value(s) used. Usually the database id will be used.

Please paste in your data below:

| id | field | value |
|-----|-----------|-------|
| 100 | serogroup | B |
| 101 | serogroup | B |

Options

Primary selection field:

Optional selection field:

☒ Update existing values

Allele designations

☒ Add additional new designation
☐ Replace existing designations

Action

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

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#)

Batch isolate update

The following changes will be made to the database. Please check that this is what you intend and then press 'Upload'. If you do not wish to make these changes, press your browser's back button.

| Transaction | id | Field | New value | Value(s) currently in database | Action |
|-------------|-----|-----------|-----------|--------------------------------|-----------------------------|
| 1 | 100 | serogroup | B | C | update field with new value |
| 2 | 101 | serogroup | B | C | update field with new value |

Action

[Back to main page](#)

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


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

| isolate | country | field | value |
|---------|---------|-----------|-------|
| CN100 | UK | serogroup | B |
| CN103 | UK | serogroup | B |

6.10 Deleting multiple isolate records


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

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



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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* PubMLST


Add, update or delete records

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


Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-----------------------------------|-----|-------------------------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attributes | + | ++ | ? | Define additional attributes to associate with values of a particular isolate record field. |

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Help Toggle: Field help: id

Isolate query/update

Isolate provenance/phenotype fields

Combine with: AND

date entered = 2014-03-17 +

curator (surname) = Jolley

Display/sort options

Order by: id ascending

Display: 25 records per page

Action

Reset Submit

3 records returned. Click the hyperlinks for detailed information.

Delete Tag scanning Projects

Delete ALL Scan Select project... Link

| Delete | Update | Sequence bin | New version | Isolate fields | | | | | | | Seqbin size (bp) | Contigs | MLST | | Finotyping antigens | | | |
|------------------------|------------------------|------------------------|------------------------|----------------|---------|---------|----------------|------|------------------------------|------------------------|------------------|---------|------|-------|----------------------------|----------|----------|---------|
| | | | | id | isolate | aliases | country | year | disease | species | serogroup | | | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| delete | update | upload | create | 28783 | 0012/14 | | Czech Republic | 2014 | carrier | Neisseria meningitidis | W | 0 | 0 | 10733 | | 5-2 | 10-1 | F5-8 |
| delete | update | upload | create | 28784 | 0014/14 | | Czech Republic | 2014 | carrier | Neisseria meningitidis | B | 0 | 0 | 35 | ST-35 complex | 22-1 | 14 | F4-1 |
| delete | update | upload | create | 28785 | 0015/14 | | Czech Republic | 2014 | invasive (unspecified/other) | Neisseria meningitidis | B | 0 | 0 | 10734 | ST-41/44 complex/Lineage 3 | 22-1 | 14 | F5-2 |

You will have a final chance to change your mind:

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Delete multiple isolate records

Warning

If you proceed, you will delete 3 isolate records. Please confirm that this is your intention.

Action

Confirm deletion!

Click 'Confirm deletion!'.

6.11 Uploading sequence contigs linked to isolate records

6.11.1 Select isolate from drop-down list

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

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

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

Database curator's interface - Neisseria PubMLST

Add, update or delete records

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

Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-------------------|-----|---------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |

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

Paste sequence contigs in FASTA format in to the form.

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

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

Batch insert sequences

This page allows you to upload sequence data for a specified isolate record in FASTA format.

If an isolate id is chosen, then all sequences will be associated with that isolate. Alternatively, the isolate id, or any other isolate table field that uniquely defines the isolate, can be named in the identifier rows of the FASTA file. This allows data for multiple isolates to be uploaded.

Please note that you can reach this page for a specific isolate by [querying isolates](#) and then clicking 'Upload' within the isolate table.

Please fill in the following fields - required fields are marked with an exclamation mark (!).

Paste in sequences in FASTA format:

```

TTTAATTCTAAATATATTTTTTCAAAGTTACTCATATTTAGTTAGTTTCCTTTTACT
GGGTAGTAGTAATCAAATTTCTGACTTATCTGTAAATACTTTAATTGAGTTGTGGGT
TGTCCACCTTCTTTAATAGAGTAGTACCAATAAATTTCTTACATCGACAGTCCGCATA
TATTGGCCATCAGGAGTCATCGATACGGGAGAGAACTACTTTTATTACTTTGAAGTATA
ACCTTCAATTCATTTGGGAGATGGTAAAGTGAACGGTTATTGCAATAGGCTATGG
AAGTGGCCCTCAAGAACATGCTCAAAGCCAGCAGATACCGGTTGTCCCGTTTGTCTATT
GGTGTATATCGAGTAGTACTATTGCTATATTAATTCGTGATTATTACAGCACCTATATT
TTTGATTCTGTAGTAGGCAATCAAGTCTTTAACTGTCTGGGTTGGTATAGCCTGCTTG
SCGCGCTTTCGCTTTCGACACGSCCCACAGCGCTTTCCAGACGCTTGGCTACCGTTACC
SCACCGGTGGCGATTCCGCGCCGCTTCCGCGAGCTGAGTGACCATGACAGTACACCA
GAGGATTAGCATGCGAGTGTGATAGCTAATTTACCGCTGTACCGATCAGCGAGCT
TCCCAACTGACGATAAACCCATAGCTGGTAAATCAAAATCGGCGCTGTGATGCCATTA
CGGATATTGCTTACCAATGGCAGCATCCTTATCCTCGGATTAGTCATCGCACTGCC
SCATGTGCGAGGCAATAATTTTGATGATTTTTCAGTGCAGTTTGTGCGGCTTCTGC
GGTGTGCTTTTTCGATTTGGTAGGGTACTGTCAAATCAAAGCATTATTCACTACC
SCCAACTCAGCGCATTCGCGCAGTATTACACATCGCGCGCTTGAGTGCCGCAAGCTG
CCGCAATAATCTTCGATTAATGATACCTTATGCTTTCCGCAATCGCTGAGTGTAGAC
GGGTTTCTGCGGCAATGATTCAGCTACGATTTCCCAACTGCTGCGCAATTGCC
CGTCTTTACATTTTCTTTGACCAATCGCTAACACACCCAGCCAAAGCGTGGGCGAAC
TGTTTGGCAACATA
  
```

Attributes

isolate id: 2) 120M

identifier field: id

sender: Jolley, Keith (keith)

method:

run id:

assembly id:

Options

☐ Don't insert sequences shorter than 250 bps.

Link to experiment:

Alternatively upload FASTA file

Select FASTA file: No file selected.

Reset Submit

Back

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?]

Batch insert sequences

The following sequences will be entered.

| Original designation | Sequence length | Comments |
|---|-----------------|----------|
| 180426 NODE_1211_length_26135_cov_16.536217 | 26167 | |
| 180427 NODE_619_length_1208_cov_9.455298 | 1240 | |
| 180428 NODE_655_length_1534_cov_15.418513 | 1566 | |
| 180429 NODE_675_length_1548_cov_17.753876 | 1580 | |
| 180430 NODE_31_length_8390_cov_14.525030 | 8422 | |
| 180431 NODE_254_length_2721_cov_18.400587 | 2753 | |
| 180432 NODE_262_length_1955_cov_8.388747 | 1987 | |
| 180433 NODE_252_length_30791_cov_15.767627 | 30823 | |
| 180434 NODE_189_length_536_cov_31.078358 | 568 | |
| 180435 NODE_716_length_511_cov_12.113503 | 543 | |
| 180436 NODE_465_length_13739_cov_15.131669 | 13771 | |
| 180437 NODE_38_length_15888_cov_17.174660 | 15920 | |
| 180438 NODE_778_length_2398_cov_8.673060 | 2430 | |
| 180439 NODE_1765_length_657_cov_8.754947 | 689 | |
| 180440 NODE_729_length_255_cov_13.007843 | 287 | |
| 180441 NODE_52_length_16337_cov_17.192997 | 16369 | |
| 180442 NODE_190_length_652_cov_116.434052 | 684 | |
| 180443 NODE_95_length_3094_cov_14.927279 | 3126 | |
| 180444 NODE_770_length_5072_cov_12.878943 | 5104 | |
| 180445 NODE_263_length_1223_cov_10.451349 | 1255 | |
| 180446 NODE_181_length_4496_cov_14.635231 | 4528 | |
| 180447 NODE_558_length_4436_cov_14.670198 | 4468 | |
| 180448 NODE_1179_length_917_cov_13.504908 | 949 | |
| 180449 NODE_527_length_4033_cov_15.608480 | 4065 | |
| 180450 NODE_212_length_2435_cov_12.223409 | 2467 | |
| 180451 NODE_54_length_790_cov_9.398734 | 822 | |
| 180452 NODE_146_length_12761_cov_14.578794 | 12793 | |
| 180453 NODE_11_length_16414_cov_13.733703 | 16446 | |
| 180454 NODE_386_length_3699_cov_13.323060 | 3731 | |
| 180455 NODE_215_length_395_cov_4.250633 | 427 | |

- Number of contigs: 359
- Minimum length: 265
- Maximum length: 33563
- Total length: 2059411
- Mean length: 5736
- N50: 10581
- N90: 3126
- N95: 1863

Upload

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?]

Batch insert sequences

This page allows you to upload sequence data for a specified isolate record in FASTA format.

If an isolate id is chosen, then all sequences will be associated with that isolate. Alternatively, the isolate id, or any other isolate table field that uniquely defines the isolate, can be named in the identifier rows of the FASTA file. This allows data for multiple isolates to be uploaded.

Please note that you can reach this page for a specific isolate by [querying isolates](#) and then clicking ‘Upload’ within the isolate table.

Please fill in the following fields - required fields are marked with an exclamation mark (!).

Paste in sequences in FASTA format:

```
>0021/93
GAAGCGAAAAAATCATTGACGAAGGCGGCTTGGTGCSCGACGACATCATTAICGSCATG
GTCAAGAAACGCAICGCGCAAGACGACTGCAAAAACGGTTTCCTGTTCGACGGTTCCCG
CGCACATTGGCACAAGCCGAAGCGATGGTGAAGCAGGCGTGGATTGGATGCAATCGTT
GAAATCGACGTGCTGACAGCGTGATGTGCGACCGGATGAGCGGCCCGCGGTGCAATTG
SCITCCGSCGTACTTACCACGTITACCTACACCCGCCCAAGTGAAGGCAAGACGAC
GTAAACGCGGAAGATTGATTGAGCGGACGACGACAAAGAGAAACCGTGAAGAAACGCG
CTTGCGGTTTACCAGGACGAAGGTTTGGTTCGATTTTACAGCAAACTGGAAGGCG
GAACACGCGCTAAATATATCAAAAGTTGACGCGCACTCAGCCGGTA
>0030/93
GAAGCGAAAAAATCATTGACGAAGGCGGCTTGGTGCSCGACGACATCATTAICGSCATG
GTCAAGAAACGCAICGCGCAAGACGACTGCAAAAACGGTTTCCTGTTCGACGGTTCCCG
CGCACATTGGCACAAGCCGAAGCGATGGTGAAGCAGGCGTGGATTGGATGCAATCGTT
GAAATCGACGTGCTGACAGCGTGATGTGCGACCGGATGAGCGGCCCGCGGTGCAATTG
SCITCCGSCGTACTTACCACGTITACCTACACCCGCCCAAGTGAAGGCAAGACGAC
GTAAACGCGGAAGATTGATTGAGCGGACGACGACAAAGAGAAACCGTGAAGAAACGCG
CTTGCGGTTTACCAGGACGAAGGTTTGGTTCGATTTTACAGCAAACTGGAAGGCG
GAACACGCGCTAAATATATCAAAAGTTGACGCGCACTCAGCCGGTA
>0033/93
GAAGCGAAAAAATCATTGACGAAGGCGGCTTGGTGCSCGACGACATCATTAICGSCATG
GTCAAGAAACGCAICGCGCAAGACGACTGCAAAAACGGTTTCCTGTTCGACGGTTCCCG
```

Attributes

isolate id: **Read identifier from FASTA**

identifier field: **isolate**

sender: Jolley, Keith (keith)

method:

run id:

assembly id:

Options

☐ Don't insert sequences shorter than 250 bps.

Link to experiment:

Alternatively upload FASTA file Action

Select FASTA file:

Browse... No file selected. **Reset** **Submit**

[Back](#)

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [Z](#)

Batch insert sequences

The following sequences will be entered. Any problems are highlighted.

| BIGSdb id | Identifier field (isolate) | Sequence length | Comments | Status |
|-----------|----------------------------|-----------------|----------|-------------|
| 938 | 0021/93 | 465 | | Will upload |
| 944 | 0030/93 | 465 | | Will upload |
| 946 | 0033/93 | 465 | | Will upload |

[Upload](#)

6.11.2 Select from isolate query


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

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


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* PubMLST

 Add, update or delete records



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

 Database configuration


| Table | Add | Batch Add | Update or delete | Comments |
|----------------------------------|-----|-----------|------------------|---|
| loci | + | ++ | ? | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attribute | + | ++ | ? | |

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Help  Toggle:  Field help: [id](#)

Isolate query/update


Isolate provenance/phenotype fields: isolate = FAM18 + 

Display/sort options: Order by: id ascending Display: 25 records per page Modify form options

Action:


1 record returned. Click the hyperlink for detailed information.

Delete Tag scanning Projects

| Delete | Update | Sequence bin | New version | Isolate fields  | | | | | | | | Seqbin size (bp) | Contigs | MLST | | Finotyping antigens | | |
|--------|--------|---------------------------------------|---------------------------------------|--|---------|-------------------|---------|------|------------------------------|------------------------|-----------|------------------|---------|------|-----------------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| delete | update | upload | create | id | isolate | aliases | country | year | disease | species | serogroup | | | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| | | | | 698 | FAM18 | NIBSC_3076; Z4259 | USA | 1983 | invasive (unspecified/other) | Neisseria meningitidis | C | 2194961 | 1 | 11 | ST-11 complex/ET-37 complex | 5 | 2 | F1-30 |
| | | <input type="button" value="update"/> | <input type="button" value="create"/> | | | | | | | | | | | | | <input type="button" value="update"/> | <input type="button" value="update"/> | <input type="button" value="update"/> |

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: 

Batch insert sequences

This page allows you to upload sequence data for a specified isolate record in FASTA format.

If an isolate id is chosen, then all sequences will be associated with that isolate. Alternatively, the isolate id, or any other isolate table field that uniquely defines the isolate, can be named in the identifier rows of the FASTA file. This allows data for multiple isolates to be uploaded.

Please note that you can reach this page for a specific isolate by [querying isolates](#) and then clicking 'Upload' within the isolate table.

Please fill in the following fields - required fields are marked with an exclamation mark (!).

Paste in sequences in FASTA format:

Attributes: isolate id: ! 698) FAM18
 sender: !
 method:
 run id:
 assembly id:

Options:
☐ Don't insert sequences shorter than 250 bps.
 Link to experiment:

Alternatively upload FASTA file: No file selected.

Action:

[Back](#)

6.11.3 Upload options

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

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

select the experiment from the dropdown list box.

6.12 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.

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

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

Database curator's interface - Neisseria PubMLST

Add, update or delete records

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

Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-------------------|-----|---------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |

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

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

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

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

Sequence tag scan

Please select the required isolate ids and loci for sequence scanning - use ctrl or shift to make multiple selections. In addition to selecting individual loci, you can choose to include all loci defined in schemes by selecting the appropriate scheme description. By default, loci are only scanned for an isolate when no allele designation has been made or sequence tagged. You can choose to rescan loci with existing designations or tags by selecting the appropriate options.

Isolates

- 19) S3131
- 24) S4355
- 30) 14
- 31) 10
- 33) 20
- 35) 26
- 46) 255 [old version]
- 52) 243
- 61) 393
- 64) 254
- 67) S5611

All None

Loci

- 16S_rDNA
- 16S_rRNA (SSU_rRNA)
- 23S_rRNA
- abcZ
- abcZ (NEIS1015)
- ackF (NEIS1279)
- ackA2 (NEIS1727)
- acnA (NEIS1729)
- acnB (NEIS1492)
- adhA (NEIS0486)
- adhC (NEIS1241)

All None

Schemes

- Capsule
- Genetic Information Processing
- Genomic Islands
- Lineage Schemes
- Metabolism
- Pilin
- Typing
- MLST
- Finotyping antigens

Parameters

Min % identity: 70

Min % alignment: 50

BLASTN word size: 20

Return up to: 1 partial match(es)

Stop after: 200 new matches

Stop after: 5 minute(s)

☐ Use TBLASTX

☐ Hunt for nearby start and stop codons

☐ Return partial matches even when exact matches are found

☒ Rescan even if allele designations are already set

☐ Rescan even if allele sequences are tagged

☐ Mark missing sequences as provisional allele '0'

Restrict included sequences by

Sequence method:

Project:

Experiment:

Action

Reset Scan

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

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

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

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

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

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

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

Sequence tag scan

Please select the required isolate ids and loci for sequence scanning - use ctrl or shift to make multiple selections. In addition to selecting individual loci, you can choose to include all loci defined in schemes by selecting the appropriate scheme description. By default, loci are only scanned for an isolate when no allele designation has been made or sequence tagged. You can choose to rescan loci with existing designations or tags by selecting the appropriate options.

Isolates

- 19) S3131
- 24) S4355
- 30) 14
- 31) 10
- 35) 26
- 46) 255 [old version]
- 52) 243
- 61) 393
- 64) 254
- 67) S5611

Loci

- *16S_rDNA
- 16S_rRNA (SSU_rRNA)
- _23S_rRNA
- abcZ (NEIS1015)
- acqF (NEIS1279)
- ackA2 (NEIS1727)
- acnA (NEIS1729)
- acnB (NEIS1492)
- adhA (NEIS0486)
- adhC (NEIS1241)

Schemes

- Capsule
- Genetic Information Processing
- Genomic Islands
- Lineage Schemes
- Metabolism
- Pilin
- Typing
- MLST
- Finotyping antigens

Parameters

Min % identity: 70
 Min % alignment: 50
 BLASTN word size: 20
 Return up to: 1 partial match(es)
 Stop after: 200 new matches
 Stop after: 5 minute(s)

☐ Use TBLASTX
☐ Hunt for nearby start and stop codons
☐ Return partial matches even when exact matches are found
☒ Rescan even if allele designations are already set
☐ Rescan even if allele sequences are tagged
☐ Mark missing sequences as provisional allele '0'

Restrict included sequences by

Sequence method:
 Project:
 Experiment:

Action

Reset Scan

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

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

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

Sequence tag scan

| Isolate | 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 → | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 34) 20 | exact | adk | 3 | 100.00 | 465 | 465 | 0.0 | 182750 | 1392 | 1856 | 1392 | 1856 | extract → | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 34) 20 | exact | aroE | 1 | 100.00 | 490 | 490 | 0.0 | 182731 | 32577 | 33066 | 32577 | 33066 | extract → | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 34) 20 | exact | tumC | 1 | 100.00 | 465 | 465 | 0.0 | 182815 | 19783 | 20247 | 19783 | 20247 | extract → | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 34) 20 | exact | gdh | 1 | 100.00 | 501 | 501 | 0.0 | 182852 | 7516 | 8016 | 7516 | 8016 | extract → | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 34) 20 | exact | pdhC | 1 | 100.00 | 480 | 480 | 0.0 | 182819 | 13868 | 14347 | 13868 | 14347 | extract → | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 34) 20 | exact | pgm | 3 | 100.00 | 450 | 450 | 0.0 | 182831 | 24559 | 25008 | 24559 | 25008 | extract → | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |

Action

Tag alleles/sequences

Started: Thu Jul 2 10:47:30 2015
 Finished: Thu Jul 2 10:47:45 2015
 Elapsed time: 15 seconds

Please note that scan results will remain on the server for 7 days.

Individual sequences can be extracted for inspection by clicking the ‘extract →’ 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

6.13 Projects

6.13.1 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.


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - Neisseria PubMLST

 Add, update or delete records

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

 Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-------------------|-----|-------------------------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |

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.


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - Neisseria PubMLST

 Add, update or delete records

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

 Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-----------------------------------|-----|---------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | ? | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | ? | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attributes | + | ++ | ? | Define additional attributes to associate with values of a particular isolate record field. |

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'.

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?](#)

Add new project member

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record


project id: Action

isolate id:

curator:


timestamp:

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



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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - Neisseria PubMLST


Add, update or delete records

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


Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-----------------------------------|-----|---------------------|------------------|---|
| loci | + | ++ databank scan | ? | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | ? | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | ? | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attributes | + | ++ | ? | Define additional attributes to associate with values of a particular isolate record field. |

Download an Excel submission template:

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?]

Batch insert project members

This page allows you to upload project member data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)

Paste in tab-delimited text (include a field header line).

Action

[Back](#)

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [?]

Batch insert project members

This page allows you to upload project member data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- [Download tab-delimited header for your spreadsheet](#) - use Paste special → text to paste the data.
- [Download submission template \(xlsx format\)](#)

Paste in tab-delimited text (include a field header line).

| project_id | isolate_id |
|------------|------------|
| 3 | 18968 |
| 3 | 18969 |
| 3 | 19023 |
| 3 | 19024 |
| 3 | 19025 |
| 3 | 19026 |
| 3 | 19027 |
| 3 | 19028 |
| 3 | 19029 |
| 3 | 19030 |
| 3 | 19031 |
| 3 | 19032 |
| 3 | 19958 |
| 3 | 19959 |
| 3 | 19960 |
| 3 | 19961 |
| 3 | 19962 |
| 3 | 19963 |
| 3 | 19964 |
| 3 | 19965 |

Action

[Back](#)

6.14 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:

The screenshot shows the 'Full information on isolate 5' page. At the top is a navigation bar with links: Query, Search, Browse, Profile/ST, List, Breakdown, Isolate fields, Scheme/alleles, Publications, Links, Contents, Home, Options, Profiles/sequences definitions, Database submissions. The main content area is divided into sections: Provenance/meta data, Versions, Publication (1), and Schemes and loci. The Provenance/meta data section contains two columns of information: id: 21, isolate: 5, strain designation: NG: P1.ND.ND: F-ND: ST-ND (-), country: Germany, continent: Europe, region: Volkach, year: 1999, age yr: 18, disease: carrier, source: throat swab, epidemiology: endemic, species: Neisseria meningitidis, serogroup: NG, comments: Bavarian carriage study, sender: Ulrich Vogel, University of Wuerzburg, Wuerzburg, Germany, curator: Keith Jolley, University of Oxford, UK (E-mail: keith.jolley@zoo.ox.ac.uk), update history: 1 update, show details, date entered: 2014-06-30, datestamp: 2014-06-30. The Versions section is highlighted with a red box and contains the text 'More than one version of this isolate record exist.' and 'Older versions: 20'. The Publication (1) section lists a citation: Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M, Vogel U (2005). Genetic analysis of meningococci carried by children and young adults. *J Infect Dis* 191:1263-71. The Schemes and loci section shows a message: 'No loci available for analysis.'


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

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


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Database curator's interface - *Neisseria* PubMLST

 Add, update or delete records

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

 Database configuration

| Table | Add | Batch Add | Update or delete | Comments |
|-----------------------------------|-----|---------------|------------------|---|
| loci | + | ++ | ? | |
| | | databank scan | | |
| locus aliases | + | ++ | ? | Add alternative names for loci. These can also be set when you batch add loci. |
| PCR reactions | + | ++ | ? | Set up <i>in silico</i> PCR reactions. These can be used to filter genomes for tagging to specific repetitive loci. |
| nucleotide probes | + | ++ | ? | Define nucleotide probes for <i>in silico</i> hybridization reaction to filter genomes for tagging to specific repetitive loci. |
| isolate field extended attributes | + | ++ | ? | Define additional attributes to associate with values of a particular isolate record field. |

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

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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) Toggle: [Field help](#): id

Browse *Neisseria* PubMLST database

— Browse criteria — Action

Order by: [id](#) Browse all records

Direction: [ascending](#)

Display: 25 records per page [Field help](#)

☐ Include old record versions

36583 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

[Delete](#)
[Tag scanning](#)
[Projects](#)


[Delete ALL](#)
[Scan](#)
[Select project...](#)
[Link](#)

Page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [Last](#)

| Delete | Update | Sequence bin | New version | id | isolate | aliases | country | year | disease | species | serogroup | ST | MLST | Finotyping antigens | | |
|------------------------|------------------------|------------------------|------------------------|----|-----------|------------------------|----------------|------|------------------------------|-------------------------------|-----------|------|----------------------------|---------------------|-----------|-------------|
| | | | | | | | | | | | | | | PorA VR1 | PorA VR2 | PorA VR |
| delete | update | upload | create | 1 | A4M1027 | B1; NIBSC_2803; Z1001 | USA | 1937 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | A | 4 | ST-4 complex/subgroup IV | 5-2 update | 10 update | F1-5 update |
| delete | update | upload | create | 2 | 120M | B35; NIBSC_2822; Z1035 | Pakistan | 1967 | meningitis and septicemia | <i>Neisseria meningitidis</i> | A | 1 | ST-1 complex/subgroup VII | 5-2 update | 10 update | F5-1 update |
| delete | update | upload | create | 3 | M00242905 | | UK | 2000 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | B | 1099 | | 19 update | 15 update | add |
| delete | update | upload | create | 4 | M1027 | B43; NIBSC_3076; Z1043 | USA | 1937 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | A | 4 | ST-4 complex/subgroup IV | add | add | add |
| delete | update | upload | create | 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | B | 1100 | ST-32 complex/ET-5 complex | 7 update | 16 update | add |
| delete | update | upload | create | 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | W | 1101 | ST-22 complex | add | add | add |
| delete | update | upload | create | 7 | 7891 | B54; NIBSC_2760; Z1054 | Finland | 1975 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | A | 5 | ST-5 complex/subgroup III | 20 update | 9 update | F3-1 update |
| delete | update | upload | create | 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | B | 1102 | ST-18 complex | add | 14 update | add |
| delete | update | upload | create | 9 | 0021/84 | | Czech Republic | 1984 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | W | 114 | ST-22 complex | add | add | add |
| delete | update | upload | create | 10 | 8748 | B73; NIBSC_3704; Z1073 | Czech Republic | 1974 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | A | 4 | ST-4 complex/subgroup IV | 19 update | 2 update | EE-1 update |

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.


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Create new isolate record version

This page allows you to create a new version of the isolate record shown below. Provenance and publication information will be copied to the new record but the sequence bin and allele designations will not. This facilitates storage of different versions of genome assemblies. The old record will be hidden by default, but can still be accessed when needed, with links from the new record. The update history will be reset for the new record.

Enter new record id

Options

Action

id: 30457
☐ Add new version to projects
 Create

Provenance/meta data

| | | |
|--|--|---|
| id: 7 | source: CSF | comments: Pili I,IIa |
| isolate: 7891 | epidemiology: epidemic | sender: Wendell Zollinger, Dept Bacterial Diseases, Walter Reed Army Institute of Research, Washington DC, USA |
| aliases: B54; Z1054 | species: <i>Neisseria meningitidis</i> | curator: Carina Brehony, University of Oxford, UK (E-mail: carina.brehony@zoo.ox.ac.uk) |
| strain designation: A: P1.20.9; F3-1; ST-5 (cc5) | serogroup: A | update history: 35 updates show details |
| country: Finland | MLEE designation: Subgroup III | date entered: 2001-02-07 |
| continent: Europe | serotype: 4,21 | datestamp: 2014-06-17 |
| year: 1975 | sero subtype: P1.9 | |
| disease: invasive (unspecified/other) | ET no: 48 | |


Publications (7) [show/hide](#)

Sequence bin

| | | |
|--------------------------|-------------------|---|
| contigs: 199 | N50: 23361 | detailed breakdown: Display |
| total length: 2057385 bp | N90: 6218 | |
| max length: 112831 bp | N95: 3616 | |
| mean length: 10339 bp | loci tagged: 1673 | |


Curating submitted data

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



[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Neisseria locus/sequence definitions database


The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.


Query database


- [Sequence query](#) - query an allele sequence.
- [Batch sequence query](#) - query multiple sequences in FASTA format.
- [Sequence attribute search](#) - find alleles by matching attributes.
- [Browse MLST profiles](#)
- [Search MLST profiles](#)
- [List](#) - find MLST profiles matched to entered list.
- [Search by combinations of MLST alleles](#) - including partial matching.
- [Batch profile query](#) - lookup MLST profiles copied from a spreadsheet.
- [Extract finetype from whole genome data](#) (experimental) **NEW**


Downloads


- [Allele sequences](#)
- [MLST profiles](#)


Option settings


- [Set general options](#)


Submissions


- [Manage submissions](#)


General information

- Number of sequences: 124772
- Number of profiles (MLST): 10026
- Last updated: 2015-06-18
- [Profile update history](#)
- [About BIGSdb](#)


Export

- [Sequences](#) - XMFA / concatenated FASTA formats


Analysis

- [Sequence similarity](#) - find sequences most similar to selected allele.
- [Sequence comparison](#) - display a comparison between two sequences.
- [Locus Explorer](#) - tool for analysing allele sequences stored for particular locus.

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- alleles
- MLST profiles

New allele sequence submissions waiting for curation

Your account is authorized to handle the following submissions:

| Submission id | Submitted | Updated | Submitter | Locus | Sequences |
|-----------------------------------|------------|------------|------------|----------|-----------|
| BIGSdb_20150623074942_31862_84622 | 2015-06-23 | 2015-06-23 | Joe Bloggs | NEIS0001 | 3 |

[Return to index page](#)

7.1 Alleles

Click the link to the appropriate submission on the ‘Manage submissions’ page.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- alleles
- MLST profiles

New allele sequence submissions waiting for curation

Your account is authorized to handle the following submissions:

| Submission id | Submitted | Updated | Submitter | Locus | Sequences |
|-----------------------------------|------------|------------|------------|----------|-----------|
| BIGSdb_20150623074942_31862_84622 | 2015-06-23 | 2015-06-23 | Joe Bloggs | NEIS0001 | 3 |

[Return to index page](#)

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

Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FeTA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623074942_31862_84622

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

timestamp: 2015-06-23

status: pending

outcome: accepted - data uploaded

locus: NEIS0001

sequences: 3

technology: Illumina

read length: 100-199

coverage: 20-49x

assembly: de novo

assembly software: Velvet

Sequences

| Identifier | Length | Sequence | Complete CDS | Status | Assigned allele |
|------------|--------|--|--------------|---------|------------------------|
| UK322 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |
| UK323 | 924 | ATGCTGGAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |
| UK347 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |

Batch curate

Update

Messages

Add message

Archive

Archive of submission and any supporting files:

Download

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

Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FeTA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623074942_31862_84622

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

timestamp: 2015-06-23

status: pending

outcome: accepted - data uploaded

locus: NEIS0001

sequences: 3

technology: Illumina

read length: 100-199

coverage: 20-49x

assembly: de novo

assembly software: Velvet

Sequences

| Identifier | Length | Sequence | Complete CDS | Status | Assigned allele |
|------------|--------|--|--------------|---------|------------------------|
| UK322 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |
| UK323 | 924 | ATGCTGGAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |
| UK347 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |

Batch curate

Update

Messages

Add message

Archive

Archive of submission and any supporting files:

Download

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.

Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FeTA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623074942_31862_84622

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-06-23

status: pending

outcome: accepted - data uploaded

locus: NEIS0001

sequences: 3

technology: Illumina

read length: 100-199

coverage: 20-49x

assembly: de novo

assembly software: Velvet

Sequences

| Identifier | Length | Sequence | Complete CDS | Status | Assigned allele |
|------------|--------|--|--------------|---------|------------------------|
| UK322 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | pending | Curate |
| UK323 | 924 | ATGCTGGAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | pending | Curate |
| UK347 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | pending | Curate |

Batch curate

Update

Messages

Add message

Archive

Archive of submission and any supporting files:

Download

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 will need to manually set the status from the dropdown list of values.

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

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

Add new allele sequence

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

locus: NEIS0001

allele id: 209

sequence: ATGCTGCAAGAAGACTTTGGC... (full sequence text)

status: [dropdown]

sender: Bloggs, Joe (jbloggs)

curator: Keith Jolley (keith)

date entered: 2015-06-23

datestamp: 2015-06-23

comments:

Flags: atypical, contains IS element, downstream fusion

Action: Reset, Submit

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#) [Toggle](#)

Add new allele sequence

Sequence NEIS0001 (lpxC): 209 added!

[Add another](#) [Return to submission](#) [Back to main page](#)

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

Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
 Download: [Alleles](#) | [MLST profiles](#)
 Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Curate submission

Submission: BIGSdb_20150623074942_31862_84622

Summary

type: alleles
 submitter: Joe Bloggs, University of Oxford, UK
 datestamp: 2015-06-23
 status: pending
 locus: NEIS0001
 sequences: 3 [FASTA](#)
 technology: Illumina
 read length: 100-199
 coverage: 20-49x
 assembly: de novo
 assembly software: Velvet

Sequences

| Identifier | Length | Sequence | Complete CDS | Status | Assigned allele |
|------------|--------|--|--------------|----------|------------------------|
| UK322 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | assigned | 209 |
| UK323 | 924 | ATGCTGGAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |
| UK347 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |

[Batch curate](#) [Update](#)

Messages

[Add message](#)

Archive

Archive of submission and any supporting files:
[Download](#) [TAR](#)

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.

Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FeTA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623074942_31862_84622

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-06-23

status: pending

locus: NEIS0001

sequences: 3

technology: Illumina

read length: 100-199

coverage: 20-49x

assembly: de novo

assembly software: Velvet

Sequences

| Identifier | Length | Sequence | Complete CDS | Status | Assigned allele |
|------------|--------|--|--------------|----------|------------------------|
| UK322 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | assigned | 209 |
| UK323 | 924 | ATGCTGGAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |
| UK347 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTTGATAA | ✓ | pending | Curate |

[Batch curate](#)

Update

Messages

[Add message](#)

Archive

Archive of submission and any supporting files:

[Download](#)

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

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

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

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

Batch insert sequences

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 (pxC)

status:

sender: Bloggs, Joe (jbloggs)

sequence (FASTA):

```
>UK323
ATGCTGGAAGAAGACTTTGGCGAAATCCATCAGCGTTACCGGAGTCGGGCTGCATTCCGGC
GAACGCGTCGGCTGACCCCTGACCCCGCGCCTGAAAACAGCGGATTTCCTTCGCCGT
ACCGATTGGACGCGAGATGGCGAACAATCAAGCTGACCCCTTATTTGATCAACGAT
ACCGCCTTTCTCCACCATCGTTACCGACAAAGGCGTGCCTCGGACGATCGAACAC
ATTATGTCGCGCTGTCCGCTACGGCATCGACAACGCGCTGATTGAGCTGAACGCGCC
GAATCCCGATTATGACGCGCTCCAGCCTGCCGTTTATTACCTTTTGCAAGATCGGGC
GTGTCGATCAAAAGGCGCAAAAGCGTTTTTTGAAAATCCTCAAGCCTGTCGAAATCAA
GAACGGGCAAAATGGGTGCGCTTTACGCCGTATGACGGCTTTAAGGTAACGCTGACCATC
GAATTCGACCATCCGTTTTCAACCGCAGCTCGCCCACTTTGAAATCGATTCGCCGC
AAATCCTACATCGACAAATCGCGCGCGCGCACTTTTCGGCTTTATGCACGAAGTGGAA
```

☒ 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.

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#) [Toggle](#)

Batch insert sequences

Sequence check

| Original designation | Allele id | Status | Action |
|----------------------|-----------|--------|--|
| UK323 | 210 | OK | Upload valid sequences |
| UK347 | 211 | OK | |

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#) [Toggle](#)

Batch insert sequences

Upload succeeded.

[Return to submission](#) [Upload more](#) | [Back to main page](#)

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FcA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623074942_31862_84622

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-06-23

status: pending

locus: NEIS0001

sequences: 3

technology: Illumina

read length: 100-199

coverage: 20-49x

assembly: de novo

assembly software: Velvet

Sequences

| Identifier | Length | Sequence | Complete CDS | Status | Assigned allele |
|------------|--------|--|--------------|----------|-----------------|
| UK322 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | assigned | 209 |
| UK323 | 924 | ATGCTGGAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | assigned | 210 |
| UK347 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | assigned | 211 |

Messages

Archive

Action

Close submission

7.1.3 Rejecting sequences

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FcA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623074942_31862_84622

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-06-23

status: pending

outcome: accepted - data uploaded

locus: NEIS0001

sequences: 3

technology: Illumina

read length: 100-199

coverage: 20-49x

assembly: de novo

assembly software: Velvet

Sequences

| Identifier | Length | Sequence | Complete CDS | Status | Assigned allele |
|------------|--------|--|--------------|----------|-----------------|
| UK322 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | rejected | Curate |
| UK323 | 924 | ATGCTGGAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | rejected | Curate |
| UK347 | 924 | ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | rejected | Curate |

Batch curate

Messages

Archive

Action

Update

7.1.4 Closing the submission

You can add a message to the submitter by entering it in the message box and clicking 'Add message'. 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 Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | Feta | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623074942_31862_84622

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-06-23

status: pending

outcome: rejected - data not uploaded

locus: NEIS0001

sequences: 3

technology: Illumina

read length: 100-199

coverage: 20-49x

assembly: de novo

assembly software: Velvet

Sequences

| Identifier | Length | Sequence | Complete CDS | Status | Assigned allele |
|------------|--------|--|--------------|----------|-----------------|
| UK322 | 924 | ATGCTGCAAAGAACITTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | rejected | |
| UK323 | 924 | ATGCTGCAAAGAACITTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | rejected | |
| UK347 | 924 | ATGCTGCAAAGAACITTTGGC ... ACATCAGAAATTGTGGATAA | ✓ | rejected | |

Update

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-23 07:44:40+00 | Keith Jolley | These sequences are not NEIS0001 alleles. |

Add message

Archive

Archive of submission and any supporting files:

Download

Action

Close submission

7.2 Profiles

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | Feta | Options | Isolate Database

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

Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- alleles
- MLST profiles

New allelic profile submissions waiting for curation

Your account is authorized to handle the following submissions:

| Submission id | Submitted | Updated | Submitter | Scheme | Profiles |
|----------------------------------|------------|------------|------------|--------|----------|
| BIGSdb_20150623105542_5928_87752 | 2015-06-23 | 2015-06-23 | Joe Bloggs | MLST | 3 |

Return to index page

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623105542_5928_87752

Summary

type: profiles


submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-06-23

status: pending

Archive

Archive of submission and any supporting files:

Download 

Profiles

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|---------|------------------------|
| UK32 | 43 | 2 | 12 | 32 | 32 | 3 | 2 | pending | Curate |
| UK33 | 7 | 56 | 4 | 3 | 2 | 12 | 12 | pending | Curate |
| UK34 | 76 | 3 | 5 | 3 | 87 | 43 | 34 | pending | Curate |

Batch curate

Messages

Add message

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 Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623105542_5928_87752

Summary

type: profiles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-06-23

status: pending

Messages

Add message


Profiles

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|---------|------------------------|
| UK32 | 43 | 2 | 12 | 32 | 32 | 3 | 2 | pending | Curate |
| UK33 | 7 | 56 | 4 | 3 | 2 | 12 | 12 | pending | Curate |
| UK34 | 76 | 3 | 5 | 3 | 87 | 43 | 34 | pending | Curate |

Batch curate

Archive

Archive of submission and any supporting files:

Download 

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#) [Toggle](#)

Add new MLST profile

Please fill in the fields below - required fields are marked with an exclamation mark (!).

Record

ST: ! 10024
 adk: ! 43
 abcZ: ! 2
 aroE: ! 12
 fumC: ! 32
 gdh: ! 32
 pdhC: ! 3
 pgm: ! 2
 sender: ! Bloggs, Joe (jbloggs)
 clonal_complex:
 curator: ! Keith Jolley (keith)
 date_entered: ! 2015-06-23
 datestamp: ! 2015-06-23
 PubMed ids:

Action

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#) [Toggle](#)

Add new MLST profile

ST-10024 added!

[Return to submission](#) [Add another](#) | [Back to main page](#)

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623105542_5928_87752

Summary

type: profiles
 submitter: Joe Bloggs, University of Oxford, UK
 datestamp: 2015-06-23
 status: pending

Profiles

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|----------|------------------------|
| UK32 | 43 | 2 | 12 | 32 | 32 | 3 | 2 | assigned | 10024 |
| UK33 | 7 | 56 | 4 | 3 | 2 | 12 | 12 | pending | Curate |
| UK34 | 76 | 3 | 5 | 3 | 87 | 43 | 34 | pending | Curate |

Batch curate


Update

Messages

Add message

Archive

Archive of submission and any supporting files:

Download 

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 Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623105542_5928_87752

Summary

type: profiles
 submitter: Joe Bloggs, University of Oxford, UK
 datestamp: 2015-06-23
 status: pending

Profiles

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|----------|------------------------|
| UK32 | 43 | 2 | 12 | 32 | 32 | 3 | 2 | assigned | 10024 |
| UK33 | 7 | 56 | 4 | 3 | 2 | 12 | 12 | pending | Curate |
| UK34 | 76 | 3 | 5 | 3 | 87 | 43 | 34 | pending | Curate |

Batch curate


Update

Messages

Add message

Archive

Archive of submission and any supporting files:

Download 

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

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#) [Toggle](#)

Batch insert MLST profiles

This page allows you to upload profiles 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.
- 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, then you must also provide it for each profile record.
- [Download tab-delimited header for your spreadsheet](#) - use Paste Special ▾ Text to paste the data.
- [Download submission template \(xlsx format\)](#)

Please paste in tab-delimited text (include a field header line)

| | | | | | | |
|-----|------|------|------|-----|------|-----|
| adk | abcZ | aroE | fumC | gdh | pdhC | pgm |
| 7 | 56 | 4 | 3 | 2 | 12 | 12 |
| 76 | 3 | 5 | 3 | 87 | 43 | 34 |

Parameters

Sender: [Bloggs, Joe \(jbloggs\)](#)
 Value will be overridden if you include a sender field in your pasted data.

☐ Ignore duplicate profiles

Action

[Reset](#) [Submit](#)

[Back](#)

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

Database: [Species home](#) | [Curator's page \(species\)](#) | [Curator's page \(database\)](#)
 Users: [Add](#) | [Query/update](#)
 Loci: [Add](#)
 MLST profiles: [Add](#) | [Query/update](#) | [Batch insert](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#) [Help](#) [Toggle](#)

Batch insert MLST profiles

Database updated ok

[Return to submission](#) [Back to main page](#)

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | Feta | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623105542_5928_87752

Summary

type: profiles
submitter: Joe Bloggs, University of Oxford, UK
datestamp: 2015-06-23
status: pending

Profiles


| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|----------|-------------|
| UK32 | 43 | 2 | 12 | 32 | 32 | 3 | 2 | assigned | 10024 |
| UK33 | 7 | 56 | 4 | 3 | 2 | 12 | 12 | assigned | 10025 |
| UK34 | 76 | 3 | 5 | 3 | 87 | 43 | 34 | assigned | 10026 |

Messages

Add message

Archive

Archive of submission and any supporting files:

Download 

Action

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 Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | Feta | Options | Isolate Database

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




Curate submission

Submission: BIGSdb_20150623105542_5928_87752

Summary

type: profiles
submitter: Joe Bloggs, University of Oxford, UK
datestamp: 2015-06-23
status: pending
outcome: accepted - data uploaded

Profiles

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|------------|--|
| UK32 | 43 | 2 | 12 | 32 | 32 | 3 | 2 | rejected ▼ |  Curate |
| UK33 | 7 | 56 | 4 | 3 | 2 | 12 | 12 | rejected ▼ |  Curate |
| UK34 | 76 | 3 | 5 | 3 | 87 | 43 | 34 | rejected ▼ |  Curate |


Batch curate

Messages

Add message

Archive

Archive of submission and any supporting files:

Download 

Action

Update

7.2.4 Closing the submission

You can add a message to the submitter by entering it in the message box and clicking 'Add message'. 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 Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | Feta | Options | Isolate Database

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

Curate submission

Submission: BIGSdb_20150623105542_5928_87752

Summary

type: profiles
 submitter: Joe Bloggs, University of Oxford, UK
 datestamp: 2015-06-23
 status: pending
 outcome: rejected - data not uploaded

Profiles

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|----------|-------------|
| UK32 | 43 | 2 | 12 | 32 | 32 | 3 | 2 | rejected | |
| UK33 | 7 | 56 | 4 | 3 | 2 | 12 | 12 | rejected | |
| UK34 | 76 | 3 | 5 | 3 | 87 | 43 | 34 | rejected | |

Update


Messages

| Timestamp | User | Message |
|------------------------|--------------|--|
| 2015-06-23 10:42:38+00 | Keith Jolley | You need to submit some representative isolate data for this profiles. |

Add message

Archive

Archive of submission and any supporting files:

Download 

Action

Close submission

7.3 Isolates

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

PubMLST Query: Search | Browse | Profile/ST | List
 Breakdown: Isolate fields | Scheme/alleles | Publications
 Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

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

Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- isolates

New isolate submissions waiting for curation

Your account is authorized to handle the following submissions:

| Submission id | Submitted | Updated | Submitter | Isolates |
|-----------------------------------|------------|------------|------------|----------|
| BIGSdb_20150623132929_11187_32555 | 2015-06-23 | 2015-06-23 | Joe Bloggs | 2 |

Return to index page

You will see a table summarizing the submission.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Curate submission

Submission: BIGSdb_20150623132929_11187_32555

Summary

type: isolates
 submitter: Joe Bloggs, University of Oxford, UK
 datestamp: 2015-06-23
 status: pending

Isolates

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK233 | UK | 2015 | meningitis and septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK322 | UK | 2014 | meningitis | blood | Neisseria meningitidis | B | 2 | 3 | 17 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2-1 |


Batch curate

Record status: pending

Messages

Archive

Archive of submission and any supporting files:

Download 

Add message

Click the 'Batch curate' button.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Curate submission

Submission: BIGSdb_20150623132929_11187_32555

Summary

type: isolates
 submitter: Joe Bloggs, University of Oxford, UK
 datestamp: 2015-06-23
 status: pending

Isolates

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK233 | UK | 2015 | meningitis and septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK322 | UK | 2014 | meningitis | blood | Neisseria meningitidis | B | 2 | 3 | 17 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2-1 |


Batch curate

Record status: pending

Messages

Archive


Archive of submission and any supporting files:


Download 

Add message

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

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


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)
Toggle: 

Batch insert isolates

This page allows you to upload isolate data as tab-delimited text or copied from a spreadsheet.

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.
- Enter references for your isolates as a semi-colon (;) separated list of PubMed ids (non-integer ids will be ignored).
- You can also upload allele fields along with the other isolate data - simply create a new column with the locus name (see the 'allowed_loci' tab in the Excel template for locus names). These will be added with a confirmed status and method set as 'manual'.
- You can choose whether or not to include an id number field - if it is omitted, the next available id will be used automatically.
- Download tab-delimited header for your spreadsheet - use 'Paste Special' > 'Text' to paste the data.
- Download submission template (xlsx format)

Please select the sender from the list below:

Value will be overridden if you include a sender field in your pasted data.

Paste in tab-delimited text (include a field header line).


| isolate | country | year | disease | source | species | serogroup | abcZ | adk |
|--------------|---------|------|------------|--------|-------------|------------------------|-----------|-----|
| aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 | |
| UK233 | UK | 2015 | meningitis | and | septicaemia | CSF | Neisseria | |
| meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 |
| F1-5 | 5 | 2 | | | | | | |
| UK322 | UK | 2014 | meningitis | | blood | Neisseria meningitidis | B | |
| 2 | 3 | 17 | 3 | 8 | 4 | 6 | F1-5 | 5 |
| | | | | | | | | 2-1 |

[Back](#)

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.


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

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)
Toggle: 


Batch insert isolates

Database updated ok

[Return to submission](#)
[Back to main page](#)

7.3.1 Closing the submission

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


Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Curate submission

Submission: BIGSdb_20150623132929_11187_32555

Summary

type: isolates
submitter: [Joe Bloggs](#), University of Oxford, UK
datestamp: 2015-06-23
status: pending

Isolates


| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK233 | UK | 2015 | meningitis and septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK322 | UK | 2014 | meningitis | blood | Neisseria meningitidis | B | 2 | 3 | 17 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2-1 |

Record status: accepted

Messages

Archive

Archive of submission and any supporting files:



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



Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Curate submission

Submission: BIGSdb_20150623132929_11187_32555

Summary

type: isolates
 submitter: [Joe Bloggs](#), University of Oxford, UK
 datestamp: 2015-06-23
 status: pending

Isolates

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK233 | UK | 2015 | meningitis and septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK322 | UK | 2014 | meningitis | blood | Neisseria meningitidis | B | 2 | 3 | 17 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2-1 |

Record status: accepted Update

Messages

Add message

Archive

Archive of submission and any supporting files:

[Download](#)

Action

Close submission

Offline curation tools

8.1 Automated offline sequence tagging

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

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

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

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

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

```
autotag.pl --database <database configuration>
```

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

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

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

A full list of options can be found by typing:

```
autotag.pl --help

NAME
    autotag.pl - BIGSdb automated allele tagger

SYNOPSIS
    autotag.pl --database NAME [options]

OPTIONS
    -0, --missing
        Marks missing loci as provisional allele 0. Sets default word size to 15.
    -d, --database NAME
```

```
Database configuration name.

-h, --help
    This help page.

-i, --isolates LIST
    Comma-separated list of isolate ids to scan (ignored if -p used).

--isolate_list_file FILE
    File containing list of isolate ids (ignored if -i or -p used).

-I, --exclude_isolates LIST
    Comma-separated list of isolate ids to ignore.

-l, --loci LIST
    Comma-separated list of loci to scan (ignored if -s used).

-L, --exclude_loci LIST
    Comma-separated list of loci to exclude

-m, --min_size SIZE
    Minimum size of seqbin (bp) - limit search to isolates with at least this
    much sequence.

-n, --new_only
    New (previously untagged) isolates only. Combine with --new_max_alleles
    if required.

--new_max_alleles ALLELES
    Set the maximum number of alleles that can be designated or sequences
    tagged before an isolate is not considered new when using the --new_only
    option.

-o, --order
    Order so that isolates last tagged the longest time ago get scanned first
    (ignored if -r used).

--only_already_tagged
    Only check loci that already have a tag present (but no allele
    designation). This must be combined with the --already_tagged option or
    no loci will match. This option is used to perform a catch-up scan where
    a curator has previously tagged sequence regions prior to alleles being
    defined, without the need to scan all missing loci.

-p, --projects LIST
    Comma-separated list of project isolates to scan.

-P, --exclude_projects LIST
    Comma-separated list of projects whose isolates will be excluded.

-q, --quiet
    Only error messages displayed.

-r, --random
    Shuffle order of isolate ids to scan.

-R, --locus_regex REGEX
    Regex for locus names.
```

```

-s, --schemes LIST
    Comma-separated list of scheme loci to scan.

-t, --time MINS
    Stop after t minutes.

--threads THREADS
    Maximum number of threads to use.

-T, --already_tagged
    Scan even when sequence tagged (no designation).

-v, --view VIEW
    Isolate database view (overrides value set in config.xml).

-w, --word_size SIZE
    BLASTN word size.

-x, --min ID
    Minimum isolate id.

-y, --max ID
    Maximum isolate id.

```

8.2 Automated offline allele definition

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

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

```

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

```

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

```

scannew.pl --database <database configuration>

```

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

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

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

A full list of options can be found by typing:

```

scannew.pl --help

NAME
    scannew.pl - BIGSdb automated allele definer

SYNOPSIS
    scannew.pl --database NAME [options]

```

```
OPTIONS
-a, --assign
    Assign new alleles in definitions database.

-A, --alignment INT
    Percentage alignment (default: 100).

-B, --identity INT
    Percentage identity (default: 99).

-c, --coding_sequences
    Only return complete coding sequences.

-d, --database NAME
    Database configuration name.

-h, --help
    This help page.

-i, --isolates LIST
    Comma-separated list of isolate ids to scan (ignored if -p used).

--isolate_list_file FILE
    File containing list of isolate ids (ignored if -i or -p used).

-I, --exclude_isolates LIST
    Comma-separated list of isolate ids to ignore.

-l, --loci LIST
    Comma-separated list of loci to scan (ignored if -s used).

-L, --exclude_loci LIST
    Comma-separated list of loci to exclude.

-m, --min_size SIZE
    Minimum size of seqbin (bp) - limit search to isolates with at least this
    much sequence.

-n, --new_only
    New (previously untagged) isolates only.

-o, --order
    Order so that isolates last tagged the longest time ago get scanned first
    (ignored if -r used).

-p, --projects LIST
    Comma-separated list of project isolates to scan.

-P, --exclude_projects LIST
    Comma-separated list of projects whose isolates will be excluded.

-r, --random
    Shuffle order of isolate ids to scan.

-R, --locus_regex REGEX
    Regex for locus names.

-s, --schemes LIST
```

```

    Comma-separated list of scheme loci to scan.

-t, --time MINS
    Stop after t minutes.

--threads THREADS
    Maximum number of threads to use.

-T, --already_tagged
    Scan even when sequence tagged (no designation).

-v, --view VIEW
    Isolate database view (overrides value set in config.xml).

-w, --word_size SIZE
    BLASTN word size.

-x, --min ID
    Minimum isolate id.

-y, --max ID
    Maximum isolate id.

```

8.3 Cleanly interrupting offline curation

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

```
kill 23232
```

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

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

```
pkill autotag
```

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

Similarly, to terminate all allele autodefiner jobs, type

```
pkill scannew
```

8.4 Uploading contigs from the command line

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

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

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

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

A full list of options can be found by typing:

```
upload_contigs.pl --help

NAME
    upload_contigs.pl - Upload contigs to BIGSdb isolate database

SYNOPSIS
    upload_contigs.pl --database NAME --isolate ID --file FILE
                      --curator ID --sender ID [options]

OPTIONS
-a, --append
    Upload contigs even if isolate already has sequences in the bin.

-c, --curator ID
    Curator id number.

-d, --database NAME
    Database configuration name.

-f, --file FILE
    Full path and filename of contig file.

-h, --help
    This help page.

-i, --isolate ID
    Isolate id of record to upload to.

-m, --method METHOD
    Method, e.g. 'Illumina', default 'unknown'.

-s, --sender ID
    Sender id number.
```

Definition downloads

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

9.1 Allele sequence definitions

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

9.1.1 Scheme tree



The screenshot shows the PubMLST website interface. At the top, there is a navigation bar with links: Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query. Below this, there are links for Download: Alleles | MLST profiles, and Links: Contents | Home | PorA | FetA | Options | Isolate Database. The main heading is 'Download allele sequences'. Below the heading, there are three tabs: 'Select loci by scheme' (highlighted with a red box), 'Alphabetical list', and 'All loci by scheme'. A text box below the tabs says: 'Click within the tree to display details of loci belonging to schemes or groups of schemes - clicking a group folder will display the loci for all schemes within the group and any subgroups. Click the nodes to expand/collapse.' Below this is a hierarchical tree structure with the following nodes: All loci, Capsule, Genetic Information Processing, Genomic islands, Metabolism, Typing, Other schemes, and Loci not in schemes.

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.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Download allele sequences

Select loci by scheme | [Alphabetical list](#) | [All loci by scheme](#)

Click within the tree to display details of loci belonging to schemes or groups of schemes - clicking a group folder will display the loci for all schemes within the group and any subgroups. Click the nodes to expand/collapse.

- All loci
 - Capsule
 - Genetic Information Processing
 - Genomic islands
 - Metabolism
 - Typing
 - MLST**
 - Finotyping antigens
 - rplF species
 - Antigen genes
 - eMLST (20 locus partial genes)
 - eMLST (20 locus whole genes)
 - N. meningitidis cgMLST v1.0
 - Other schemes
 - Loci not in schemes

MLST

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

Download table: [tab-delimited text](#) | [Excel format](#)

Click the download link for the required locus

MLST

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

Download table: [tab-delimited text](#) | [Excel format](#)

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

```
>fumC_1
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCCGCTGCGGT
TTGGGCGAAATCAAAATCCCCGAAAACGAGCCGGGTTCTGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTTCGGCAACGACGTT
ACCATCGGTATGGCGGGCGCGTCGGGCAATTTTCGAGCTGAACGTCTATATGCCCGTCATC
```


```

GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCGTCGGCATTGAACCCGTACCGGAAAAAATCGACTATTTCTGCACCATTC
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAAC
>fumC_2
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCCGCGCTGCGGT
TTGGGCGAAATCAAATCCCCGAAAACGAGCCGGGTTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGCATGGCGGGCGCGTCGGGCAATTTTCGAGCTGAACGTCTATATGCCCGTTATC
GCCTACAACCTCTTGCAATCCATCCGCCTCTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCATCGGCATCGAACCCGTACCGGAAAAAATCGACTATTTCTGCACCATTC
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAAC
>fumC_3
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCCGCGCTGCGGT
TTGGGCGAAATCAAATCCCCGAAAACGAGCCGGGTTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGCATGGCGGGCGCGTCGGGCAATTTTCGAGCTGAACGTCTATATGCCCGTTATC
GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCGTCGGCATCGAACCCGTACCGGAAAAAATCGACTATTTCTGCACCATTC
CTGATGCTGGTTACTGCGTTAAACCGTAAAATCGGCTACGAAAAAC

```

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.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

[Help](#)

Download allele sequences

[Select loci by scheme](#) | [Alphabetical list](#) | [All loci by scheme](#)

A

| Locus | Download | Type | Alleles | Length | Full name/product | Curator(s) | Last updated |
|------------------|----------|------|---------|--------------------------------|---|------------------------|--------------|
| ArsR [NEIS1769] | | DNA | 70 | Variable: No limits set | ArsR family transcriptional regulator | | 2015-03-18 |
| AsnC [NEIS1566] | | DNA | 160 | Variable: No limits set | transcription regulator AsnC | | 2015-05-20 |
| abcZ | | DNA | 728 | Fixed: 433 bp | | O. Harrison, K. Jolley | 2015-06-22 |
| abcZ [NEIS1015] | | DNA | 493 | Variable: No limits set | ABC transporter ATP-binding protein | | 2015-05-19 |
| aceF [NEIS1279] | | DNA | 477 | Variable: (1563 min; 1641 max) | dihydrolipoamide acetyltransferase (EC 2.3.1.12) | | 2015-05-19 |
| ackA2 [NEIS1727] | | DNA | 408 | Variable: No limits set | acetate kinase | | 2015-05-20 |
| acnA [NEIS1729] | | DNA | 527 | Variable: No limits set | aconitate hydratase 1 (EC 4.2.1.3) | | 2015-05-20 |
| acnB [NEIS1492] | | DNA | 476 | Variable: No limits set | aconitate hydratase 2 (EC 4.2.1.3) | | 2015-05-19 |
| adhA [NEIS0486] | | DNA | 614 | Variable: No limits set | alcohol dehydrogenase | | 2015-05-17 |
| adhC [NEIS1241] | | DNA | 239 | Variable: No limits set | alcohol dehydrogenase | | 2015-05-20 |
| adk | | DNA | 501 | Fixed: 465 bp | | O. Harrison, K. Jolley | 2015-06-08 |
| adk [NEIS0767] | | DNA | 208 | Variable: No limits set | adenylate kinase | | 2015-05-18 |
| aldA [NEIS1942] | | DNA | 447 | Variable: No limits set | aldehyde dehydrogenase | | 2015-05-20 |
| aniA [NEIS1549] | | DNA | 0 | Variable: No limits set | nitrite reductase, major outer membrane copper-containing protein | | |
| anmK [NEIS1788] | | DNA | 480 | Variable: No limits set | anhydro-N-acetylmuramic acid kinase | A. Jamet | 2015-05-20 |
| apaH [NEIS0610] | | DNA | 116 | Variable: No limits set | diadenosine tetraphosphatase | | 2015-05-17 |
| argH [NEIS0580] | | DNA | 433 | Variable: No limits set | argininosuccinate lyase | | 2015-05-18 |
| aroE | | DNA | 775 | Fixed: 490 bp | | O. Harrison, K. Jolley | 2015-06-22 |
| aroE [NEIS1810] | | DNA | 305 | Variable: No limits set | shikimate dehydrogenase | | 2015-05-21 |
| aspA | | DNA | 189 | Fixed: 432 bp | | K. Jolley | 2015-04-21 |
| aspA [NEIS1185] | | DNA | 501 | Variable: No limits set | aspartate ammonia-lyase | | 2015-05-20 |
| atIA [NEIS2274] | | DNA | 18 | Variable: No limits set | atIA / peptidoglycan transglycosylase | O. Harrison | 2014-12-10 |
| autA [NEIS1859] | | DNA | 242 | Variable: No limits set | autotransporter A | | 2015-05-20 |

Click the download links for the required locus.

9.1.3 All loci by scheme

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | Feta | Options | Isolate Database

[Help](#)

Download allele sequences

Select loci by scheme | [Alphabetical list](#) | All loci by scheme

MLST

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

Finetyping antigens

| Locus | Download | Type | Alleles | Length | Full name/product | Curator(s) | Last updated |
|----------|--------------------------|---------|---------|-------------------------|------------------------|------------|--------------|
| PorA VR1 | Download | peptide | 264 | Variable: No limits set | PorA variable region 1 | K. Jolley | 2015-06-16 |
| PorA VR2 | Download | peptide | 735 | Variable: No limits set | PorA variable region 2 | K. Jolley | 2015-06-16 |
| Feta VR | Download | peptide | 581 | Variable: No limits set | | I. Feavers | 2015-06-26 |

rpIF species

| Locus | Download | Type | Alleles | Length | Full name/product | Curator(s) | Last updated |
|-------|--------------------------|------|---------|---------------|------------------------------------|------------|--------------|
| 'rpIF | Download | DNA | 123 | Fixed: 413 bp | 50S ribosomal protein L6 (partial) | E. Watkins | 2015-01-30 |

ADP-heptose biosynthesis

| Locus | Download | Type | Alleles | Length | Full name/product | Aliases | Curator(s) | Last updated |
|-----------------|--------------------------|------|---------|-------------------------|---|------------------------|------------|--------------|
| NEIS0769 (hldA) | Download | DNA | 244 | Variable: No limits set | D-beta-D-heptose-7-phosphate kinase | NMB0825; NMC0769; rfaE | C. Kahler | 2015-05-18 |
| NEIS0773 (hldD) | Download | DNA | 311 | Variable: No limits set | ADP-D-beta-D heptose epimerase | NMB0828; NMC0773 | C. Kahler | 2015-05-17 |
| NEIS2014 (gmhB) | Download | DNA | 216 | Variable: No limits set | D-alpha,beta,D-Heptose 1,7 bisphosphate phosphatase | NMB2033; NMC2014 | C. Kahler | 2015-05-20 |
| NEIS2055 (hldC) | Download | DNA | 206 | Variable: No limits set | D-beta-D-heptose-1-phosphate adenyllyltransferase | NMB2076; NMC2055 | C. Kahler | 2015-05-21 |
| NEIS2070 (gmhA) | Download | DNA | 230 | Variable: No limits set | sedoheptulose-7-phosphate isomerase | NMB2090; NMC2070 | C. Kahler | 2015-05-21 |

Click the green download links for the required locus.

9.1.4 Download locus table

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

Genetic Information Processing
Genomic islands
Metabolism
Typing
MLST
Finetyping antigens
rpIF species
Antigen genes
eMLST (20 locus partial genes)
eMLST (20 locus whole genes)
N. meningitidis cgMLST v1.0
Other schemes
Loci not in schemes

MLST

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

Download table [tab-delimited text](#) | [Excel format](#)

9.2 Scheme profile definitions

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | Options | PubMLST.org | Isolate Database

Campylobacter locus/sequence definitions database

The Campylobacter PubMLST sequence definition database contains allele and profile data representing the total known diversity of *C. jejuni* and *C. coli*. Every new ST deposited in this database should have a corresponding record in the isolate database.

Query database

- Sequence query - query an allele sequence.
- Batch sequence query - query multiple sequences in FASTA format.
- Sequence attribute search - find alleles by matching attributes.
- Browse MLST profiles
- Search MLST profiles
- List - find MLST profiles matched to entered list.
- Search by combinations of MLST alleles - including partial matching.
- Batch profile query - lookup MLST profiles copied from a spreadsheet.

Downloads

- Allele sequences
- MLST profiles**

Option settings

- Set general options

Submissions

- Manage submissions

General information

- Number of sequences: 412175
- Number of profiles (MLST): 7841
- Last updated: 2015-06-17
- Profile update history
- About BIGSdb

Export

- Sequences - XMFA / concatenated FASTA formats

Analysis

- Sequence similarity - find sequences most similar to selected allele.
- Sequence comparison - display a comparison between two sequences.
- Locus Explorer - tool for analysing allele sequences stored for particular locus.

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Toggle: ⓘ

Neisseria locus/sequence definitions database

The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.

Query database

- Sequence query - query an allele sequence.
- Batch sequence query - query multiple sequences in FASTA format.
- Sequence attribute search - find alleles by matching attributes.
- Browse profiles
- Search profiles
- List - find profiles matched to entered list.
- Search by combinations of alleles - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

Downloads

- Allele sequences
- MLST** Download profiles

Option settings

- Set general options

Submissions

- Manage submissions

General information

- Number of sequences: 519864
- Number of profiles: [Show](#)
- Last updated: 2015-06-26
- Profile update history
- About BIGSdb

Export

- Sequences - XMFA / concatenated FASTA formats

Analysis

- Sequence similarity - find sequences most similar to selected allele.
- Sequence comparison - display a comparison between two sequences.
- Locus Explorer - tool for analysing allele sequences stored for particular locus.

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

| | | | | | | | | |
|----|------|-----|------|------|-----|------|-----|----------------------------|
| ST | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | clonal_complex |
| 1 | 1 | 3 | 1 | 1 | 1 | 1 | 3 | ST-1 complex/subgroup I/II |

| | | | | | | | | |
|----|---|----|----|----|---|----|----|-----------------------------|
| 2 | 1 | 3 | 4 | 7 | 1 | 1 | 3 | ST-1 complex/subgroup I/II |
| 3 | 1 | 3 | 1 | 1 | 1 | 23 | 13 | ST-1 complex/subgroup I/II |
| 4 | 1 | 3 | 3 | 1 | 4 | 2 | 3 | ST-4 complex/subgroup IV |
| 5 | 1 | 1 | 2 | 1 | 3 | 2 | 3 | ST-5 complex/subgroup III |
| 6 | 1 | 1 | 2 | 1 | 3 | 2 | 11 | ST-5 complex/subgroup III |
| 7 | 1 | 1 | 2 | 1 | 3 | 2 | 19 | ST-5 complex/subgroup III |
| 8 | 2 | 3 | 7 | 2 | 8 | 5 | 2 | ST-8 complex/Cluster A4 |
| 9 | 2 | 3 | 8 | 10 | 8 | 5 | 2 | ST-8 complex/Cluster A4 |
| 10 | 2 | 3 | 4 | 2 | 8 | 15 | 2 | ST-8 complex/Cluster A4 |
| 11 | 2 | 3 | 4 | 3 | 8 | 4 | 6 | ST-11 complex/ET-37 complex |
| 12 | 4 | 3 | 2 | 16 | 8 | 11 | 20 | |
| 13 | 4 | 10 | 15 | 7 | 8 | 11 | 1 | ST-269 complex |
| 14 | 4 | 1 | 15 | 7 | 8 | 11 | 1 | ST-269 complex |

Data records

Record pages for different types of data can be accessed following a query by clicking appropriate hyperlinks.

10.1 Isolate records

An Isolate record page displays everything known about an isolate.



[Query: Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
[Breakdown: Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
[Links: Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Full information on isolate M10 240474

Projects

This isolate is a member of the following projects:

MRF Meningococcus Genome Library

The [MRF Meningococcus Genome Library](#) is a collaboration between Public Health England, 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.



899_MRF

MRF-MGL isolates epi years 2010/2011 to 2011/2012 excluding Northern Ireland

E&W_genogroup_B_MRF

All MRF Meningococcus Genome Library genogroup B isolates from England and Wales in 2010/11 and 2011/12

MRF_no_NI

All MRF genomes excluding Northern Ireland

Provenance/meta data

| | |
|---|---|
| id: 18968 | species: <i>Neisseria meningitidis</i> |
| isolate: M10 240474 | serogroup: B |
| strain designation: B: P1.19-1,15-11: F3-9: ST-269 (cc269) | ENA accession: ERR086224 → www.ebi.ac.uk |
| country: UK | sender: Dorothea Hill |
| continent: Europe | curator: Dorothea Hill, University of Oxford, UK (E-mail: dorothea.hill@zoo.ox.ac.uk) |
| region: South East | update history: 103 updates show details |
| year: 2010 | date entered: 2012-02-15 |
| epidemiological year: 07/2010-06/2011 | datestamp: 2014-06-23 |
| disease: invasive (unspecified/other) | |

Sequence bin

| | |
|---------------------------------|--|
| contigs: 275 | N90: 6405 |
| total length: 2195045 bp | N95: 3513 |
| max length: 109859 bp | loci tagged: 1611 |
| mean length: 7982 bp | detailed breakdown: Display |
| N50: 34308 | |

Schemes and loci



Navigate and select schemes within tree to display allele designations

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

10.1.1 Projects

Projects

This isolate is a member of the following projects:

MRF Meningococcus Genome Library

The [MRF Meningococcus Genome Library](#) is a collaboration between Public Health England, 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.



This displays a list of projects that the isolate is a member of. Only projects that have a full description will be displayed.

10.1.2 Provenance metadata

Provenance/meta data

| | | | |
|------------------------------|--|------------------------|---|
| id: | 18968 | species: | Neisseria meningitidis |
| isolate: | M10 240474 | serogroup: | B |
| strain designation: | B: P1.19-1,15-11: F3-9: ST-269 (cc269) | ENA accession: | ERR086224 → www.ebi.ac.uk |
| country: | UK | sender: | Dorothea Hill |
| continent: | Europe | curator: | Dorothea Hill, University of Oxford, UK (E-mail: dorothea.hill@zoo.ox.ac.uk) |
| region: | South East | update history: | 103 updates show details |
| year: | 2010 | date entered: | 2012-02-15 |
| epidemiological year: | 07/2010-06/2011 | timestamp: | 2014-06-23 |
| disease: | invasive (unspecified/other) | | |

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.

[Back to isolate information](#)

| Timestamp | Curator | Action |
|------------------|----------------|---|
| 2014-06-23 09:44 | Dorothea Hill | NEIS2363: new designation '28' (sequence bin scan) NEIS2363: sequence tagged. Seqbin id: 4836; 3070-5108 (sequence bin scan) |
| 2014-06-17 10:55 | Carina Brehony | gyrA: designation '1000' deleted |
| 2014-06-16 16:17 | Carina Brehony | gyrA: sequence tagged. Seqbin id: 4885; 23385-23909 (sequence bin scan) |
| 2014-06-16 14:48 | Carina Brehony | gyrA: new designation '12' (sequence bin scan) |
| 2014-06-05 15:55 | Carina Brehony | gyrA: new designation '1000' |
| 2014-06-05 15:55 | Carina Brehony | gyrA: designations '12,30' deleted |
| 2014-05-30 13:26 | Eloise Orton | gyrA: new designation '30' (sequence bin scan) gyrA: sequence tagged. Seqbin id: 4989; 25611-26135 (sequence bin scan) |
| 2014-05-17 09:26 | Auto Tagger | NEIS2155: new designation '244' (sequence bin scan) NEIS2155: sequence tagged. Seqbin id: 5030; 977-1984 (sequence bin scan) pip: new designation '127' (sequence bin scan) pip: sequence tagged. Seqbin id: 4818; 35321-35736 (sequence bin scan) serC: new designation '123' (sequence bin scan) serC: sequence tagged. Seqbin id: 4765; 3629-4078 (sequence bin scan) |
| 2014-03-16 11:15 | Auto Tagger | NEIS0020: new designation '3' (sequence bin scan) NEIS0020: sequence tagged. Seqbin id: 4845; 4412-5980 (sequence bin scan) NEIS0036: new designation '8' (sequence bin scan) NEIS0036: sequence tagged. Seqbin id: 4977; 2774-3817 (sequence bin scan) NEIS0408: new designation '70' (sequence bin scan) NEIS0408: sequence tagged. Seqbin id: 4850; 28381-30666 (sequence bin scan) NEIS0409: new designation '25' (sequence bin scan) NEIS0409: sequence tagged. Seqbin id: 4850; 30685-31230 (sequence bin scan) NEIS0410: new designation '7' (sequence bin scan) NEIS0410: sequence tagged. Seqbin id: 4850; 31248-31895 (sequence bin scan) NEIS0411: new designation '2' (sequence bin scan) NEIS0411: sequence tagged. Seqbin id: 4850; 31896-32495 (sequence bin scan) NEIS0412: new designation '3' (sequence bin scan) NEIS0412: sequence tagged. Seqbin id: 4850; 32498-33613 (sequence bin scan) NEIS0721: new designation '14' (sequence bin scan) NEIS0721: sequence tagged. Seqbin id: 4950; 33102-34214 (sequence bin scan) NEIS1838: new designation '12' (sequence bin scan) |

- Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007). Species status of *Neisseria gonorrhoeae*: evolutionary and epidemiological inferences from multilocus sequence typing. *BMC Biol* **5**:35 576 isolates
- Didelot X, Urwin R, Maiden MC, Falush D (2009). Genealogical typing of *Neisseria meningitidis*. *Microbiology* **155**:3176-86 93 isolates
- Jolley KA, Sun L, Moxon ER, Maiden MC (2004). Dam inactivation in *Neisseria meningitidis*: prevalence among diverse hyperinvasive lineages. *BMC Microbiol* **4**:34 84 isolates
- Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005). The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*. *Mol Biol Evol* **22**:562-9 378 isolates
- Maiden MC, Bygraves JA, Fell E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**:3140-5 107 isolates
- Thompson EA, Feavers IM, Maiden MC (2003). Antigenic diversity of meningococcal enterobactin receptor FetA, a vaccine component. *Microbiology* **149**:1849-58 107 isolates
- Urwin R, Russell JE, Thompson EA, Holmes EC, Feavers IM, Maiden MC (2004). Distribution of surface protein variants among hyperinvasive meningococci: implications for vaccine design. *Infect Immun* **72**:595-62 78 isolates
- Wang JF, Caugant DA, Li X, Hu X, Poolman JT, Crowe BA, Achtman M (1992). Clonal and antigenic analysis of serogroup A *Neisseria meningitidis* with particular reference to epidemiological features of epidemic meningitis in the People's Republic of China. *Infect Immun* **60**:5267-82 47 isolates

10.1.4 Sequence bin summary

Sequence bin

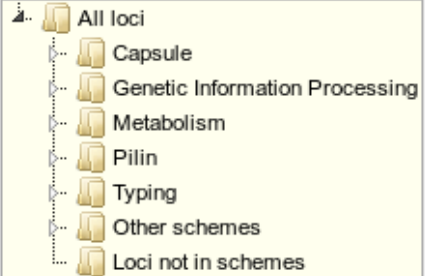
| | | | |
|---------------|------------|---------------------|--------------------------|
| contigs: | 275 | N90: | 6405 |
| total length: | 2195045 bp | N95: | 3513 |
| max length: | 109859 bp | loci tagged: | 1611 |
| mean length: | 7982 bp | detailed breakdown: | <button>Display</button> |
| N50: | 34308 | | |

This section contains basic statistics describing the sequence bin. Clicking the ‘Display’ button navigates to the *sequence bin record*.

10.1.5 Scheme and locus data

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

Schemes and loci



Navigate and select schemes within tree to display allele designations

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

Schemes and loci



MLST

| | | | | | | | | |
|---------------------|----------------------|----------------------|---------------------|---------------------|----------------------|---------------------|-----|----------------|
| abcZ | adk | aroE | fumC | gdh | pdhC | pgm | ST | clonal complex |
| 4 S | 10 S | 15 S | 9 S | 8 S | 11 S | 9 S | 269 | ST-269 complex |

Click an allele number within the scheme profile, will display the appropriate *allele definition record*. Clicking the green ‘S’ link will display the appropriate *sequence tag record*.

10.2 Allele definition records

An allele definition record displays information about a defined allele in a sequence definition database.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Allele information - abcZ: 2

Provenance/meta data

locus: [abcZ](#) 

allele: 2

sequences: TTGTATACCG TTGCGAAGG TTGGGTGAA ATTGCGATT TATTGCGCG TTACCACCG GTGCGCATG AGTGGAAAA CGGTTCGGGT GAGGCTTGT TGAAGAAT CAACGAATTA CAACTTGAAA TCGAAGCGAA
GGACGGCTGG AAGCTGGATG CGGCAGTCAA GCAGACTTGG GGGGAACG GTTTCGCGA AAACGAAAA ATCGGCAACC TTCCGCGCG TCAGAAAAAG CGTGTGCGT TGGCGCAGG TTGGGTGCG AAGCCCGGACG
TATTGCTGCT GGACGAACCG ACCAACCATT TGGATATCGA CCGGATTATT TGGTGGAAA ATCTGCTCAA AGCGTTTGA GGCAGCTTGG TTGTGATTAC CCACGACCGC CGTTTTTTGG ACAATATCG CACGCGGATT
GTGGAACG ATC

length: 433

status: Sanger trace checked

date entered: 2001-02-07

date stamp: 2009-11-11

sender: Keith Jolley, University of Oxford, UK

curator: Man-Suen Chan, University of Oxford (E-mail: man-suen.chan@paediatrics.ox.ac.uk)

Profiles containing this allele

MLST: [993 profiles](#)


Isolate databases

PubMLST isolates: Contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample. [5041 isolates](#)

If the allele is a member of a scheme profile, e.g. MLST, this will be listed. In this case, there will be a button to display all profiles of that scheme that contain the allele.

Similarly, if a *client database* has been setup for the database and the allele has been identified in an isolate, there will be a button to display all isolates that have that allele.

10.3 Sequence tag records


[Query: Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
[Breakdown: Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
[Links: Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

NEIS0346 allele sequence: id-18968

Contig position

sequence bin id: 4956 length: 744

contig length: 22993 orientation: forward

start: 13864 complete: yes

end: 14607 method: Illumina

Sequence

```

CACTGTACGC GGCACAGGC TTCAGATTG CGGCGAGCG GAAAACTAT TACGCTACG CGACGCTAA AACCGAAGT GCCGCTTAA TGGAGAAAA ATGTTAAGCG CGGCTACCT CCACCTGCAC GAAGCCTTGG GTTTGGGCC
GATGTGGTTG AAGCGGGGAG CGGTGCTCT CGGCTCGCAG ACATTGCGCG AAAGCCCGAC ACAATTCGC CGCAGAAAGC AAACCGCTCT CAGCATTCCG CAGCGTCGCT CGAACAGCA TACCGTTCAG GCACGGCTCA AAACCATGAA
AGTGTGGAAA ACACCGCGCG TACATACGCG CAACCCCGCG CTGTGAACCG AAAGCCCTCT GTCCGCGGTT TCAGACGCGA TCGCCCGCGT TCCGCGCGCT TCGGGCATAA CCAAACTTGC CGTTGTGAGC ATGTGTCCG CGACCGAGGA
TATGGTTTAC GGGCAACTGT TCCACGAGAA AGCGGGTGTG CTGCTGACAC ATATACTCAA AGCGGTAGGG CTGAGTGCGC CCTATGTCCA CAAAACCTGT TGGGTGAAAA CGCGCGCGT CGSCAACCG ATGCCGTCTG AAGCGGCAAT
CGCAATGCG CTGGACAAA CGGCCGCGCA ACTCGACGCG TGCCGCGCAC CGCGCTCT GTTCTCGGA CAGCGCTTGG TCAACCGGSA ACGGCAGCG ATGATTGAAA CTTTGTGCGC CGGCCGCCC TTCTTCATCA TCGACCATCC
CGCCCGCTG TTGCGCAAC CCGAATCAA AGCCCGCGCC TGGCAGACG TCAAGCAGCT CAAACGTGTC TTGGCGCAAG CGCGCGCGAG TTGAAGCGCG CGCAGCGGG CGTAGAATC GCAACTGCT CCAATATCT GACAGAAAGC
ACAAATGAC CGATTTCGC CAAGATTTC TAAATTCTC CTC

```

Translation

```

H C T R R T A S A L R A G G K T I T V Q P T V K P K M P S W R K Y F1
T V R G A R L Q H C G Q A E K L L P Y S R R N R R C R L N G E N F2
L Y A A H G F S I A G R R K N Y Y R T A D G K T E D A V L M E K I F3
1 CACTGTACGC GGCACAGGC TTCAGATTG CGGCGAGCG GAAAACTAT TACGCTACG CGACGCTAA AACCGAAGT GCCGCTTAA TGGAGAAAA ATGTTAAGCG CGGCTACCT CCACCTGCAC GAAGCCTTGG GTTTGGGCC 100
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
V K R A L P P P A R S L G F G P D V V E A G S R R P A V R N I A G F1
M L S A R Y L H L H E A L G L G P M W L K R G A V V L P S A T L P E F2
C A R A T S T C T K P W V W A R C G S G E P S S C R P Q H C R F3
101 ATGTTAAGCG GGCACAGGC TTCAGATTG CGGCGAGCG GAAAACTAT TACGCTACG CGACGCTAA AACCGAAGT GCCGCTTAA TGGAGAAAA ATGTTAAGCG CGGCTACCT CCACCTGCAC GAAGCCTTGG GTTTGGGCC 200
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
K P D T N P P A K A N R P Q H S A A S V R T A Y R S G T A Q N H E F1
S P T Q I R P Q K Q T V L S I P Q R P S E Q H T G Q A R L K T M K F2
K A R H K S A R K S K P S S A F R S V R P N S I P V R H G S K P K F3
201 AAAGCCCGCACCAAAATCCGCGCGCAAAAGCAACCGCTCTCAGCATTCCGCGAGCTCCGTCGCAACGATACCGGTAGGCAAGCTCAAAACCATGAA 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

Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Profile information for ST-11 (MLST)

| ST | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | clonal complex |
|----|------|-----|------|------|-----|------|-----|-----------------------------|
| 11 | 2 | 3 | 4 | 3 | 8 | 4 | 6 | ST-11 complex/ET-37 complex |

sender: Paula Kriz, Paula Kriz and Keith Jolley
curator: Keith Jolley, University of Oxford, UK (E-mail: keith.jolley@zoo.ox.ac.uk)
date entered: 2001-02-07
datestamp: 2009-11-11

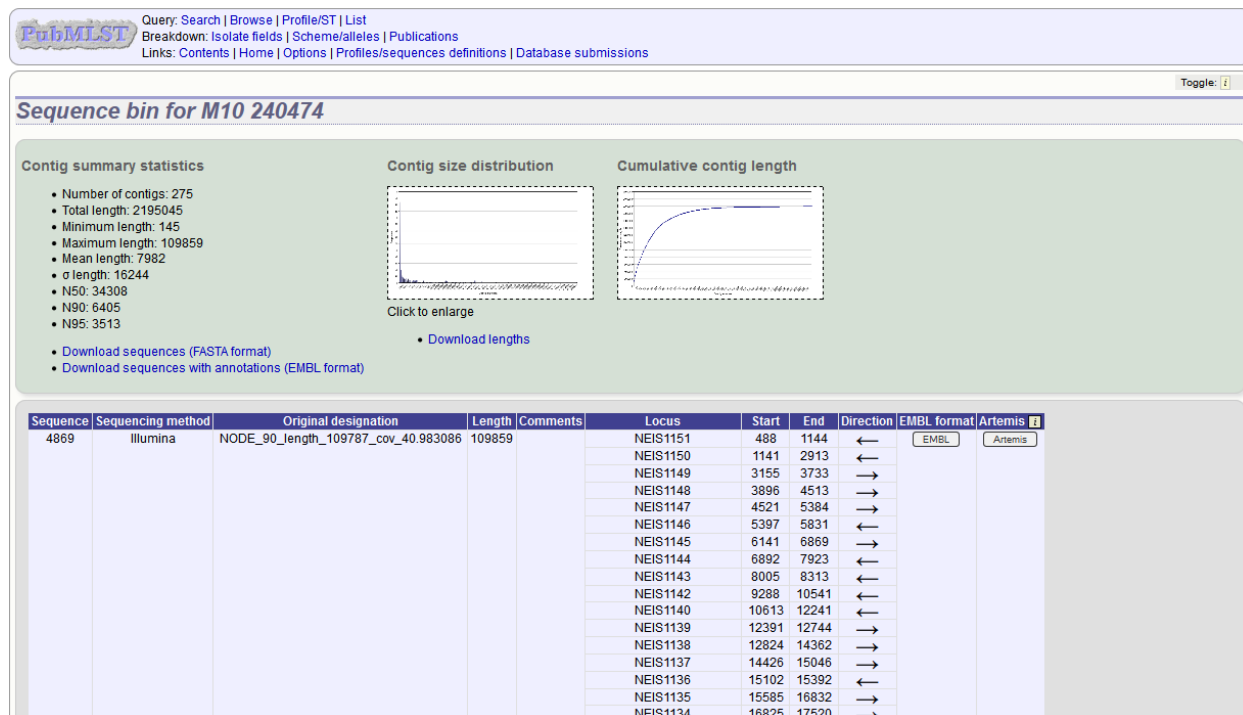
Client database

PubMLST isolates: Contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample. [2699 isolates](#)

A profile record displays information about a scheme, e.g. MLST, profile. Each allele number within the profile will be hyperlinked. Clicking these will take you to the appropriate *allele definition record*.

If a *client database* has been setup for the database and an isolate has the profile, there will be a button to display all isolates that have the profile.

10.5 Sequence bin records



A sequence bin record contains information about that contigs associated with an isolate record. This includes:

- Number of contigs
- Total length
- Minimum length
- Maximum length
- N50, N90 and N95 values
- Size distribution charts

There are also links to download the contigs in FASTA or EMBL format.

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

Querying data

11.1 Querying sequences to determine allele identity

Sequence queries are performed in the sequence definition database. Click ‘Sequence query’ from the contents page.

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Neisseria locus/sequence definitions database

The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.

Query database

- Sequence query** - query an allele sequence.
- Batch sequence query** - query multiple sequences in FASTA format.
- Sequence attribute search** - find alleles by matching attributes.
- Browse profiles**
- Search profiles**
- List** - find profiles matched to entered list.
- Search by combinations of alleles** - including partial matching.
- Batch profile query** - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data**

Downloads

- Allele sequences**
- MLST**

Option settings

- Set general options**

Submissions

- Manage submissions**

General information

- Number of sequences:** 519870
- Number of profiles:** [Show](#)
- Last updated:** 2015-07-02
- [Profile update history](#)
- [About BIGSdb](#)

Export

- Sequences** - XMFA / concatenated FASTA formats

Analysis

- Sequence similarity** - find sequences most similar to selected allele.
- Sequence comparison** - display a comparison between two sequences.
- Locus Explorer** - tool for analysing allele sequences stored for particular locus.

Paste your sequence in to the box - there is no need to trim. Normally, you can leave the locus setting on ‘All loci’ - the software should identify the correct locus based on your sequence. Sometimes, it may be quicker, however, to select the specific locus or scheme (e.g. MLST) that a locus belongs to.

Note: If the locus you are querying is a shorter version of another, e.g. an MLST fragment of a gene where the full length gene is also defined, you will need to select the specific locus or the scheme from the dropdown box. Leaving the selection on ‘All loci’ will return a match to the longer sequence in preference to the shorter one.

Click ‘Submit’.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Sequence query - *Neisseria* locus/sequence definitions

Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences. [?]

— Please select locus/scheme — Order results by —
All loci locus

Enter query sequence (single or multiple contigs up to whole genome in size)

```
GACGCGGTGCGCGATGAAAAAGTCAAAGTCATCAAGTCATTGAAGCCGCTGACCGTGGAA
TCTGTCAATGAAAAATGTCGTGCGCGGACAAATATACCGCCGCCAAAGGSCATGAACGSCAT
CTTGAAGAAATCAACTATCGGTTTGGCCAAACGACATCACGCGAGGTCAAAAACATTGCCAT
TGAGGGCAAAACCAATTTTGGCTTTTGGGCGCGGGCGGCGCGGTGCGCGCGCTAATTCGGT
TTTGAAGGACACCGCCCTGCCCGTATCGTTCGTTGCCAACCGCACCCACGCCAAAGCCGA
AGAATTGCGCGCGCTTTTCGGCATTGAAGCCGTCGCCGATGCGCGATGTGAACGCGGTTT
TGATATCATCATCA
```

Alternatively upload FASTA file

Select FASTA file:
Browse... No file selected.

Action
Reset Submit

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Sequence query - *Neisseria* locus/sequence definitions

Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The 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 locus

Enter query sequence (single or multiple contigs up to whole genome in size)

```
GACGCGGTGCGCGATGAAAAAGTCAAAGTCATCAAGTCATTGAAGCCGCTGACCGTGGAA
TCTGTCAATGAAAAATGTCGTGCGCGGACAAATATACCGCCGCCAAAGGSCATGAACGSCAT
CTTGAAGAAATCAACTATCGGTTTGGCCAAACGACATCACGCGAGGTCAAAAACATTGCCAT
TGAGGGCAAAACCAATTTTGGCTTTTGGGCGCGGGCGGCGCGGTGCGCGCGCTAATTCGGT
TTTGAAGGACACCGCCCTGCCCGTATCGTTCGTTGCCAACCGCACCCACGCCAAAGCCGA
AGAATTGCGCGCGCTTTTCGGCATTGAAGCCGTCGCCGATGCGCGATGTGAACGCGGTTT
TGATATCATCATCAACGSCAGCTCTGCGCGCTTGAGCGGTGAGCTTCCTGCCGTCAGTCC
```

Alternatively upload FASTA file

Select FASTA file:
Browse... No file selected.

Action
Reset Submit

1 exact match found.

Translate query

| Allele | Length | Start position | End position | Flags | Comments |
|---------|--------|----------------|--------------|-------|----------|
| aroE: 8 | 490 | 136 | 625 | | |

If only a partial match is found, the most similar allele is identified along with any nucleotide differences. The varying nucleotide positions are numbered both relative to the pasted in sequence and to the reference sequence. The start position of the locus within your sequence is also indicated.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Sequence query - *Neisseria* locus/sequence definitions

Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The 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 locus

Enter query sequence (single or multiple contigs up to whole genome in size)

```
GACGCGGTSCGCGATGAAAAAGTCAAAGTCATCAGTCATTGAAGCCGCTGACCGTCGAA
TCTGTCAATGAAAAATGTCGTGCGCGGACAAATACCGCCGCAAGGCATGAACGSCAT
CTTGAGAAATCAACTATCAGTTTGCCCAACGACATCACGCGAGTCAAAAACATTGCCAT
TGAGGGCAAAACCAATTTTGCTTTTGGGCGCGGGCGCGCGGTGCGCGCGCTAATTCGGT
TTTGAAGGACACCGCCCTGCGCGTATCGTTCGTTGCCAACCGCACCCACGCCAAAGCCGA
AGAATTGGCGCGGCTTTTCGGCATTGAAGCCGTCGCCGATGGCGGATGTGAACGGCGGTTT
TGATATCATCATCAACGGCAGCTCTGGCGGTAGAGCGGTGAGCTTCTGCGGTCACTCC
```

Alternatively upload FASTA file
Select FASTA file:
[Browse...](#) No file selected.

Action
[Reset](#) [Submit](#)

[Translate query](#)
Closest match: [aroE](#): 8
[Show alignment](#)

Differences
2 differences found. [?](#)

```
258 T → 393 A
333 A → 406 G
```

The locus start point is at position 136 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.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Sequence query - *Neisseria* locus/sequence definitions

Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences. [?](#)

— Please select locus/scheme — Order results by —
All loci locus


Enter query sequence (single or multiple contigs up to whole genome in size)

Alternatively upload FASTA file
Select FASTA file:
[Browse...](#) No file selected.

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.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Sequence query - *Neisseria* locus/sequence definitions

Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The 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: MLST Order results by: locus

Enter query sequence (single or multiple contigs up to whole genome in size)

```
>4758|NODE_192_length_1326_cov_47.828808
TAGAACACAGCAATATTCAAAGATTATCTGAAGTCCGAGATTCTAGATTCCCGCTTTC
GCGGGAAATGACGAAAAGCAAGCCGTAGGTCGGATACCTTGTATCCGACAAAAGCCTGCCAT
CTCAAAATAGCCGTCGGATTCCGAGAAATCCGACCTGCCAAAACCGGGCCGCGACGCTCCGGCC
GGCAGTTAGTACGCAATTCGAACAGAAACATCACAAAAAGCCCGATTCCGGATTTTCCAAAT
CGGGCTTTTTGCGCCCGTTTGTATCCCGTGAAATATCCGCATGACAAAAATATAGTG
AATTACAAAAATCAGGACAAGGCGACGAAGCCGACAGACAGTACAGATAGTACGGTAAGG
```

Alternatively upload FASTA file

Select FASTA file: Browse... No file selected.

Action: Reset Submit

7 exact matches found.

| Allele | Length | Start position | End position | Flags | Comments |
|----------|--------|----------------|--------------|-------|----------|
| abcZ: 4 | 433 | 203051 | 203483 | | |
| adk: 10 | 465 | 938327 | 938791 | | |
| aroE: 15 | 490 | 1775325 | 1775814 | | |
| fumC: 9 | 465 | 1134240 | 1134704 | | |
| gdh: 8 | 501 | 961439 | 961939 | | |
| pdhC: 11 | 480 | 1341678 | 1342157 | | |
| pgm: 9 | 450 | 1416246 | 1416695 | | |


MLST


ST 269

clonal complex ST-269 complex

11.2 Searching for specific allele definitions


To retrieve specific allele designations, click ‘Sequence attribute search’ on a sequence definition database contents page.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)


Toggle: 


Neisseria locus/sequence definitions database


The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.


Query database


- [Sequence query](#) - query an allele sequence.
- [Batch sequence query](#) - query multiple sequences in FASTA format.
- [Sequence attribute search](#) - find alleles by matching attributes.
- [Browse profiles](#)
- [Search profiles](#)
- [List](#) - find profiles matched to entered list.
- [Search by combinations of alleles](#) - including partial matching.
- [Batch profile query](#) - lookup profiles copied from a spreadsheet.
- [Extract finetype from whole genome data](#)


Downloads


- [Allele sequences](#)
- MLST  [Download profiles](#)


Option settings


- [Set general options](#)


Submissions


- [Manage submissions](#)


General information

- Number of sequences: 519870
- Number of profiles: [Show](#)
- Last updated: 2015-07-02
- [Profile update history](#)
- [About BIGSdb](#)


Export


- [Sequences](#) - XMFA / concatenated FASTA formats



Analysis

- [Sequence similarity](#) - find sequences most similar to selected allele.
- [Sequence comparison](#) - display a comparison between two sequences.
- [Locus Explorer](#) - tool for analysing allele sequences stored for particular locus.

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

Designations can be queried using *standard operators*.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Toggle: 



Query sequences for Neisseria locus/sequence definitions database

Some loci have additional fields which are not searchable from this general page. Search for these at the [locus-specific query](#) page. Use this page also for access to the sequence analysis or export plugins.

Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a locus that uses integer allele ids using the drop-down list.


Please enter your search criteria below (or leave blank and submit to return all records). Matching sequences will be returned and you will then be able to update their display and query settings.

Search criteria

locus
=
abcZ



Display

Order by: locus ascending


Display: 25 records per page


Filter query by

Action

Reset
Submit

Click submit.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Toggle: [?](#)

Query sequences for *Neisseria* locus/sequence definitions database

Some loci have additional fields which are not searchable from this general page. Search for these at the [locus-specific 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

Combine searches with: AND

locus = abcZ + [?](#)

allele id = 5

Filter query by

Action

Reset Submit

Display


Order by: locus ascending

Display: 25 records per page [?](#)

1 record returned. Click the hyperlink for detailed information.

| locus | allele id | sequence | sequence length | comments | flags |
|-------|-----------|-------------------------------------|-----------------|----------|-------|
| abcZ | 5 | TTTGATACCGTIGCC ... TCGTGAACCTCGATC | 433 | | |

Click the hyperlinked results to display allele records.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Toggle: [?](#)

Query sequences for *Neisseria* locus/sequence definitions database

Some loci have additional fields which are not searchable from this general page. Search for these at the [locus-specific query](#) page. Use this page also for access to the sequence analysis or export plugins.

Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a locus that uses integer allele ids using the drop-down list.

Please enter your search criteria below (or leave blank and submit to return all records). Matching sequences will be returned and you will then be able to update their display and query settings.

Search criteria

Combine searches with: AND

locus = abcZ + [?](#)

allele id = 5

Filter query by

Action

Reset Submit

Display

Order by: locus ascending

Display: 25 records per page [?](#)

1 record returned. Click the hyperlink for detailed information.

| locus | allele id | sequence | sequence length | comments | flags |
|-------|-------------------|-------------------------------------|-----------------|----------|-------|
| abcZ | 5 | TTTGATACCGTIGCC ... TCGTGAACCTCGATC | 433 | | |

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Allele information - *abcZ*: 5

Provenance/meta data

locus: *abcZ*

allele: 5

sequences: TTGATACCG TTGCGAAGG TTGGGGCGAA ATTGCGGATT TATTGCGCG TTATCATCAT GTGAGCCATG AGTTGGAAAA TGGTTCGAGT GAGGCTTTGT TGAAAGAGCT TAACGAATTG
CAACTTGAAA TCGAAGCGAA GGACGGCTGG AAGCTGGATG CGSCAGTCAA GCAGACTTTG GGTGAACCTTG GTTTCGAGA AAACGAAAAA ATCGGCACCC TCTCGGCGG ACAGAAAAAG
CGTGTGCCC TAGCGCAGGC TTGGGTGCAG AAGCTGATG TATTGCTGCT GGACGAACCG ACCAACCATT TGGACATTGA CGCGATTATT TGGCTGGAAA ATCTGCTTAA AGCGTTTGAA
GGACGCTGG TTGTATTAC CCACGACGC CGTTTTTGG ACAATATCG CACGCGCATC GTGGAACCTG ATC

length: 433

status: Sanger trace checked

date entered: 2001-02-07

datestamp: 2009-11-11

sender: Keith Jolley, University of Oxford, UK

curator: Man-Suen Chan, University of Oxford (E-mail: man-suen.chan@paediatrics.ox.ac.uk)

Profiles containing this allele

MLST: 183 profiles

Isolate databases

PubMLST isolates: Contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.
399 isolates

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

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Toggle:

Query sequences for *Neisseria* locus/sequence definitions database

Some loci have additional fields which are not searchable from this general page. Search for these at the [locus-specific query](#) page. Use this page also for access to the sequence analysis or export plugins.

Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a locus that uses integer allele ids using the drop-down list.

Please enter your search criteria below (or leave blank and submit to return all records). Matching sequences will be returned and you will then be able to update their display and query settings.

Search criteria

allele id

Display

Order by: locus

Display: 25

Filter query by

locus: *abcZ*

status:

sender:

curator:

allele flag:

Action

9 records returned. Click the hyperlinks for detailed information.

| locus | allele id | sequence | sequence length | comments | flags |
|-------------|-----------|--------------------------------------|-----------------|----------|-------|
| <i>abcZ</i> | 1 | TTTGATACTGTTGCC ... TTGCGAACTCGATC | 433 | | |
| <i>abcZ</i> | 2 | TTTGATAACCGTTGCC ... TTGCGAACTCGATC | 433 | | |
| <i>abcZ</i> | 3 | TTTGATAACCGTTGCC ... TTGCGAACTTGACC | 433 | | |
| <i>abcZ</i> | 4 | TTTGATAACCGTTGCC ... TTGCGAACTCGATC | 433 | | |
| <i>abcZ</i> | 5 | TTTGATAACCGTTGCC ... TCGTGAACCTCGATC | 433 | | |
| <i>abcZ</i> | 6 | TTTGATAACCGTTGCC ... TTGCGAACTCGATC | 433 | | |
| <i>abcZ</i> | 7 | TTTGATACTGTTGCC ... TTGCGAACTCGATC | 433 | | |
| <i>abcZ</i> | 8 | TTTGATAACCGTTGCC ... TTGCGAACTTGACC | 433 | | |
| <i>abcZ</i> | 9 | TTTGATAACCGTTGCC ... TTGCGAACTCGATC | 433 | | |

11.2.1 Locus-specific attributes

Some loci have *extended attribute fields*. To query these, you need to navigate to a locus-specific query page, by clicking the ‘locus-specific query’ link on the sequence attribute search page.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Toggle: [?](#)

Query sequences for *Neisseria* locus/sequence definitions database

Some loci have additional fields which are not searchable from this general page. Search for these at the **locus-specific query** page. Use this page also for access to the sequence analysis or export plugins.

Also note that some loci in this database have allele ids defined as text strings. Queries using the '<' or '>' modifiers will work alphabetically rather than numerically unless you filter your search to a locus that uses integer allele ids using the drop-down list.

Please enter your search criteria below (or leave blank and submit to return all records). Matching sequences will be returned and you will then be able to update their display and query settings.

Search criteria: locus = + [?](#)

Display: Order by: locus ascending Display: 25 records per page [?](#)

Filter query by: Action:

Pick the required locus from the dropdown box.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Toggle: [?](#)

Query PorA VR2 sequences - *Neisseria* locus/sequence definitions database

Locus: PorA_VR2 Page will reload when changed

- Further information is available for this locus.

Please enter your search criteria below (or leave blank and submit to return all records).

Locus fields: allele id = + [?](#)

Display: Order by: allele id ascending Display: 25 records per page [?](#)

Filter query by: status: allele flag:

Action:

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

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Toggle: [?]

Query PorA VR2 sequences - Neisseria locus/sequence definitions database

Locus: **PorA_VR2** Page will reload when changed

- Further information is available for this locus.

Please enter your search criteria below (or leave blank and submit to return all records).

Locus fields
 family = 2 + [?]

Display: Order by: allele id ascending status: [?]
 Display: 25 records per page [?] allele flag: [?]

Action
 Reset Submit

83 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: [1] [2] [3] [4] [5] [Last]

| locus | allele id | sequence | sequence length | comments | family | variant | old name | mAb1 | mAb1 reactivity | mAb2 | mAb2 reactivity | flags |
|----------|-----------|-----------------|-----------------|----------|--------|---------|------------|-----------|-----------------|-------|-----------------|-------|
| PorA VR2 | 2 | HFVQQIPKSPQTLVP | 15 | | 2 | | | MN16C13F4 | + | AF202 | - | |
| PorA VR2 | 2-1 | HFVQQPPKSPQTLVP | 15 | | 2 | 1 | 2b | MN16C13F4 | - | AF202 | + | |
| PorA VR2 | 2-10 | HFVQQAPKSPQTLVP | 15 | | 2 | 10 | | | | | | |
| PorA VR2 | 2-11 | HFVLQIPKSPQTLVP | 15 | | 2 | 11 | | | | | | |
| PorA VR2 | 2-12 | HFVQQIPKSPQTLVP | 15 | | 2 | 12 | | | | | | |
| PorA VR2 | 2-13 | YFVQQIPKSPQTLVP | 15 | | 2 | 13 | | MN16C13F4 | + | | | |
| PorA VR2 | 2-14 | HFVQQKLASKPTLVP | 15 | | 2 | 14 | 33 | | | | | |
| PorA VR2 | 2-15 | HFVQQKSTSKPTLVP | 15 | | 2 | 15 | 33a (33-1) | | | | | |
| PorA VR2 | 2-16 | HFVQQKPTSKPTLVP | 15 | | 2 | 16 | 33b (33-2) | | | | | |
| PorA VR2 | 2-17 | HFVQQPTSEPTLVP | 15 | | 2 | 17 | 33c (33-3) | | | | | |
| PorA VR2 | 2-18 | HFVQQIPKSPQTLVP | 15 | | 2 | 18 | | | | | | |
| PorA VR2 | 2-19 | HFVQQISQSPQTLVP | 15 | | 2 | 19 | | | | | | |

11.3 Browsing scheme profile definitions

If a sequence definition database has schemes defined that include a primary key field, i.e. collections of loci that together create profiles, e.g. for MLST, click the link to 'Browse profiles'.

[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Toggle: ⓘ

Neisseria locus/sequence definitions database

The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.

Query database

- [Sequence query](#) - query an allele sequence.
- [Batch sequence query](#) - query multiple sequences in FASTA format.
- [Sequence attribute search](#) - find alleles by matching attributes.
- [Browse profiles](#)**
- [Search profiles](#)
- [List](#) - find profiles matched to entered list.
- [Search by combinations of alleles](#) - including partial matching.
- [Batch profile query](#) - lookup profiles copied from a spreadsheet.
- [Extract finetype from whole genome data](#)

Downloads

- [Allele sequences](#)
- [MLST](#)

Option settings

- [Set general options](#)

Submissions

- [Manage submissions](#)

General information

- Number of sequences: 519870
- Number of profiles: [Show](#)
- Last updated: 2015-07-02
- [Profile update history](#)
- [About BIGSdb](#)

Export

- [Sequences](#) - XMFA / concatenated FASTA formats

Analysis

- [Sequence similarity](#) - find sequences most similar to selected allele.
- [Sequence comparison](#) - display a comparison between two sequences.
- [Locus Explorer](#) - tool for analysing allele sequences stored for particular locus.

Choose the field to order the results by, the number of results per page to display, and click 'Browse all records'.

[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [Options](#) | [PubMLST.org](#) | [Isolate Database](#)

Toggle: ⓘ

Browse profiles - Campylobacter locus/sequence definitions

Browse criteria
Order by: [ST](#)
Direction: [ascending](#)
Display: [25](#) records per page ⓘ

Action

7281 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [>](#) [Last](#)

| ST | aspA | glnA | gltA | glyA | pgm | tkf | uncA | clonal complex |
|----|------|------|------|------|-----|-----|------|----------------|
| 1 | 2 | 1 | 54 | 3 | 4 | 1 | 5 | ST-21 complex |
| 2 | 4 | 7 | 51 | 4 | 1 | 7 | 1 | ST-45 complex |
| 3 | 3 | 2 | 5 | 10 | 11 | 11 | 6 | ST-49 complex |
| 4 | 10 | 11 | 16 | 7 | 10 | 5 | 7 | ST-403 complex |
| 5 | 7 | 2 | 5 | 2 | 10 | 3 | 6 | ST-353 complex |
| 6 | 63 | 34 | 27 | 33 | 45 | 5 | 7 | |
| 7 | 8 | 10 | 2 | 2 | 14 | 12 | 6 | ST-354 complex |
| 8 | 2 | 1 | 1 | 3 | 2 | 1 | 6 | ST-21 complex |
| 9 | 1 | 6 | 22 | 24 | 12 | 7 | 1 | ST-508 complex |
| 10 | 2 | 59 | 4 | 38 | 17 | 12 | 5 | ST-433 complex |
| 11 | 48 | 7 | 10 | 4 | 1 | 7 | 1 | ST-45 complex |
| 12 | 7 | 1 | 5 | 2 | 11 | 3 | 6 | ST-353 complex |
| 13 | 2 | 1 | 1 | 9 | 2 | 1 | 5 | ST-21 complex |

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

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [PubMLST.org](#) | [Isolate Database](#)

Profile information for ST-8 (MLST)

| | | | | | | | | |
|----|------|------|------|------|-----|-----|------|----------------|
| ST | aspA | glnA | gltA | glyA | pgm | tki | uncA | clonal complex |
| 8 | 2 | 1 | 1 | 3 | 2 | 1 | 6 | ST-21 complex |

sender: Frances Colles
curator: Frances Colles, University of Oxford, UK (E-mail: frances.colles@medawar.ox.ac.uk)
date entered: 2001-05-02
datestamp: 2001-05-02

Client database

PubMLST isolates: Contains data for a collection of isolates that represent the total known diversity of *C. jejuni* and *C. coli*. 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.

49 isolates

11.4 Querying scheme profile definitions

click the link to 'Search profiles' for the appropriate scheme on the main contents page.

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [PubMLST.org](#) | [Isolate Database](#)

Campylobacter locus/sequence definitions database

The Campylobacter PubMLST sequence definition database contains allele and profile data representing the total known diversity of *C. jejuni* and *C. coli*. Every new ST deposited in this database should have a corresponding record in the isolate database.

Query database

- Sequence query - query an allele sequence.
- Batch sequence query - query multiple sequences in FASTA format.
- Sequence attribute search - find alleles by matching attributes.
- Browse MLST profiles
- Search MLST profiles**
- List - find MLST profiles matched to entered list.
- Search by combinations of MLST alleles - including partial matching.
- Batch profile query - lookup MLST profiles copied from a spreadsheet.

Downloads

- Allele sequences
- MLST profiles

Option settings

- Set general options

Submissions

- Manage submissions

General information

- Number of sequences: 412177
- Number of profiles (MLST): 7841
- Last updated: 2015-07-01
- Profile update history
- About BIGSdb

Export

- Sequences - XMFA / concatenated FASTA formats

Analysis

- Sequence similarity - find sequences most similar to selected allele.
- Sequence comparison - display a comparison between two sequences.
- Locus Explorer - tool for analysing allele sequences stored for particular locus.

Enter the search criteria you wish to search on. You may also see some drop-down list boxes that allow further filtering of results. You can add search criteria by clicking the '+' button in the 'Locus/scheme fields' section. These can be combined using 'AND' or 'OR'.

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
 Download: [Alleles](#) | [MLST profiles](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [PubMLST.org](#) | [Isolate Database](#)

Toggle: [?]

Search profiles - *Campylobacter* locus/sequence definitions

Locus/scheme fields

Combine searches with: AND

date entered > 2014-02-01 + [?]

sender (surname) = Cody

Display/sort options

Order by: ST ascending

Display: 25 records per page [?]

Filter query by

clonal complex: [?]

Action: [Reset] [Submit]

17 records returned. Click the hyperlinks for detailed information.

| ST | aspA | glnA | gltA | glyA | pgm | tkt | uncA | clonal complex |
|------|------|------|------|------|-----|-----|------|----------------|
| 7117 | 33 | 66 | 65 | 192 | 189 | 44 | 17 | |
| 7162 | 1 | 2 | 3 | 539 | 5 | 9 | 3 | ST-42 complex |
| 7172 | 8 | 1 | 6 | 3 | 2 | 1 | 384 | ST-21 complex |
| 7176 | 7 | 448 | 5 | 2 | 10 | 3 | 6 | ST-353 complex |
| 7177 | 55 | 21 | 2 | 566 | 11 | 37 | 3 | |
| 7178 | 7 | 17 | 5 | 2 | 705 | 3 | 6 | ST-353 complex |
| 7179 | 3 | 495 | 5 | 10 | 11 | 68 | 6 | ST-49 complex |
| 7180 | 79 | 496 | 420 | 372 | 20 | 1 | 17 | |
| 7181 | 121 | 497 | 421 | 567 | 706 | 560 | 94 | |
| 7182 | 7 | 2 | 5 | 72 | 22 | 561 | 6 | |

Each field can be queried using *standard operators*.

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

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
 Download: [Alleles](#) | [MLST profiles](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [PubMLST.org](#) | [Isolate Database](#)

Toggle: [?]

Search profiles - *Campylobacter* locus/sequence definitions

Locus/scheme fields

Combine searches with: AND

date entered > 2014-02-01 + [?]

sender (surname) = Cody

Display/sort options

Order by: ST ascending

Display: 25 records per page [?]

Filter query by

clonal complex: [?]

Action: [Reset] [Submit]

17 records returned. Click the hyperlinks for detailed information.

| ST | aspA | glnA | gltA | glyA | pgm | tkt | uncA | clonal complex |
|------|------|------|------|------|-----|-----|------|----------------|
| 7117 | 33 | 66 | 65 | 192 | 189 | 44 | 17 | |
| 7162 | 1 | 2 | 3 | 539 | 5 | 9 | 3 | ST-42 complex |
| 7172 | 8 | 1 | 6 | 3 | 2 | 1 | 384 | ST-21 complex |
| 7176 | 7 | 448 | 5 | 2 | 10 | 3 | 6 | ST-353 complex |
| 7177 | 55 | 21 | 2 | 566 | 11 | 37 | 3 | |
| 7178 | 7 | 17 | 5 | 2 | 705 | 3 | 6 | ST-353 complex |
| 7179 | 3 | 495 | 5 | 10 | 11 | 68 | 6 | ST-49 complex |
| 7180 | 79 | 496 | 420 | 372 | 20 | 1 | 17 | |
| 7181 | 121 | 497 | 421 | 567 | 706 | 560 | 94 | |
| 7182 | 7 | 2 | 5 | 72 | 22 | 561 | 6 | |

11.5 Investigating allele differences

11.5.1 Sequence similarity

To find sequences most similar to a selected allele within a sequence definition database, click ‘Sequence similarity’ on the contents page.

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Toggle:

Neisseria locus/sequence definitions database

The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.

Query database

- [Sequence query](#) - query an allele sequence.
- [Batch sequence query](#) - query multiple sequences in FASTA format.
- [Sequence attribute search](#) - find alleles by matching attributes.
- [Browse profiles](#)
- [Search profiles](#)
- [List](#) - find profiles matched to entered list.
- [Search by combinations of alleles](#) - including partial matching.
- [Batch profile query](#) - lookup profiles copied from a spreadsheet.
- [Extract finetype from whole genome data](#)

Downloads

- [Allele sequences](#)
- [MLST](#)

Option settings

- [Set general options](#)

Submissions

- [Manage submissions](#)

General information

- Number of sequences: 519870
- Number of profiles: [Show](#)
- Last updated: 2015-07-02
- [Profile update history](#)
- [About BIGSdb](#)

Export

- [Sequences](#) - XMFA / concatenated FASTA formats

Analysis

- [Sequence similarity](#) - find sequences most similar to selected allele.
- [Sequence comparison](#) - display a comparison between two sequences.
- [Locus Explorer](#) - tool for analysing allele sequences stored for particular locus.

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.

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Find most similar alleles - Neisseria locus/sequence definitions

This page allows you to find the most similar sequences to a selected allele using BLAST.

Select parameters

Locus:

Allele:

Number of results:

Action

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

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Find most similar alleles - *Neisseria* locus/sequence definitions

This page allows you to find the most similar sequences to a selected allele using BLAST.

— **Select parameters** — **Action** —

Locus:

Allele:

Number of results:

abcZ-5

| Allele | % Identity | Mismatches | Gaps | Alignment | Compare |
|-----------|------------|------------|------|-----------|--|
| abcZ: 657 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 657"/> |
| abcZ: 453 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 453"/> |
| abcZ: 405 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 405"/> |
| abcZ: 404 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 404"/> |
| abcZ: 213 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 213"/> |
| abcZ: 166 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 166"/> |
| abcZ: 114 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 114"/> |
| abcZ: 103 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 103"/> |
| abcZ: 646 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 646"/> |
| abcZ: 671 | 99.77 | 1 | 0 | 433/433 | <input type="button" value="Compare abcZ: 671"/> |

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

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Allele sequence comparison - *Neisseria* locus/sequence definitions

This tool allows you to select two alleles and highlight the nucleotide differences between them.

— **Select parameters** — **Action** —

Locus:

Allele #1:

Allele #2:

Nucleotide differences between abcZ: 5 and abcZ: 453


Identity: 99.77 %


[Show alignment](#)

Differences: 1
300: G → A

11.5.2 Sequence comparison


To directly compare two sequences click ‘Sequence comparison’ from the contents page of a sequence definition database.


Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)


Toggle: 

Neisseria locus/sequence definitions database


The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.


Query database


- [Sequence query](#) - query an allele sequence.
- [Batch sequence query](#) - query multiple sequences in FASTA format.
- [Sequence attribute search](#) - find alleles by matching attributes.
- [Browse profiles](#)
- [Search profiles](#)
- [List](#) - find profiles matched to entered list.
- [Search by combinations of alleles](#) - including partial matching.
- [Batch profile query](#) - lookup profiles copied from a spreadsheet.
- [Extract finetype from whole genome data](#)


Downloads


- [Allele sequences](#)
- [MLST](#)


Option settings


- [Set general options](#)


Submissions


- [Manage submissions](#)


General information

- Number of sequences: 519870
- Number of profiles: [Show](#)
- Last updated: 2015-07-02
- [Profile update history](#)
- [About BIGSdb](#)



Export

- [Sequences](#) - XMFA / concatenated FASTA formats


Analysis

- [Sequence similarity](#) - find sequences most similar to selected allele.
- [Sequence comparison](#) - display a comparison between two sequences.
- [Locus Explorer](#) - tool for analysing allele sequences stored for particular locus.

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


Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Allele sequence comparison - Neisseria locus/sequence definitions

This tool allows you to select two alleles and highlight the nucleotide differences between them.

[Select parameters](#)

[Action](#)

Locus:

Allele #1:

Allele #2:

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

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
 Download: [Alleles](#) | [MLST profiles](#)
 Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Allele sequence comparison - *Neisseria* locus/sequence definitions

This tool allows you to select two alleles and highlight the nucleotide differences between them.

Select parameters Action

Locus:

Allele #1:

Allele #2:

Nucleotide differences between abcZ: 5 and abcZ: 8

Identity: 90.53 %

[Show alignment](#)

Differences: 41

72: G → T
 78: A → G
 79: A → C
 81: T → C
 82: G → A
 83: G → A
 87: G → A
 88: A → G
 89: G → A
 90: T → C
 93: G → C
 95: C → T
 99: G → A
 102: G → A

See also:

Locus explorer plugin.

11.6 Querying isolate data

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

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The *Neisseria* PubMLST database contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database** - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [i](#) Field help: [id](#) [Go](#)

Search *Neisseria* PubMLST database

Isolate provenance/phenotype fields

Combine with: **AND**

country USA [+](#) [i](#)

year 1999

Display/sort options

Order by: **id** [i](#) ascending [v](#)

Display: **25** [i](#) records per page [i](#)

Action

[Reset](#) [Submit](#)

300 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [>](#) [Last](#)

| Isolate fields i | | | | | | | MLST | | Finotyping antigens | | | |
|----------------------------------|------------|---------|---------|------|------------------------------|------------------------|-----------|------|-----------------------------|----------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 341 | M7085 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | | | |
| 499 | MDL01A0601 | | USA | 2001 | meningitis | Neisseria meningitidis | Y | 1378 | ST-23 complex/Cluster A3 | | | |
| 500 | MDL01A2447 | | USA | 2001 | invasive (unspecified/other) | Neisseria meningitidis | Y | 1379 | ST-23 complex/Cluster A3 | | | |
| 866 | MDO1227 | | USA | 2001 | | Neisseria meningitidis | | 1624 | ST-167 complex | | | |
| 867 | MDO1056 | | USA | 2001 | | Neisseria meningitidis | | 1625 | ST-23 complex/Cluster A3 | | | |
| 868 | MDO1066 | | USA | 2001 | | Neisseria meningitidis | | 1626 | ST-269 complex | | | |
| 2281 | M7089 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | |
| 2299 | M7257 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | |
| 2316 | M7086 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 22 | ST-22 complex | 6 | 3 | |
| 2317 | M7084 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 22 | ST-22 complex | 6 | 3 | |
| 2322 | M7092 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1065 | ST-22 complex | 6 | 3 | |
| 2323 | M7100 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1286 | ST-22 complex | 6 | 3 | |
| 2324 | M7259 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1066 | ST-22 complex | 6 | 3 | |

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:

- 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.
- Filters
 - Various filters may be available, including
 - * Publications
 - * Projects
 - * MLST profile completion status
 - * Clonal complex
 - * Sequence bin size
 - * Inclusion/exclusion of *old versions*

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Toggle: ☐ Field help: id

Search *Neisseria* PubMLST database

Isolate provenance/phenotype fields

Combine with: AND

country = USA +

year > 1999

Action

Display/sort options

Order by: id

Display: 25 records per page

300 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page:

☒ Modify form options

Click to add or remove additional query terms:

- Allele designations/scheme field values
- Allele designation status
- Tagged sequence status
- Filters

| Isolate fields <input type="button" value="i"/> | | | | | | | | MLST | | Finotyping antigens | | |
|---|------------|---------|---------|------|------------------------------|------------------------|-----------|------|-----------------------------|---------------------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 341 | M7085 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | | | |
| 499 | MDL01A0601 | | USA | 2001 | meningitis | Neisseria meningitidis | Y | 1378 | ST-23 complex/Cluster A3 | | | |
| 500 | MDL01A2447 | | USA | 2001 | invasive (unspecified/other) | Neisseria meningitidis | Y | 1379 | ST-23 complex/Cluster A3 | | | |
| 866 | MDO1227 | | USA | 2001 | | Neisseria meningitidis | | 1624 | ST-167 complex | | | |
| 867 | MDO1056 | | USA | 2001 | | Neisseria meningitidis | | 1625 | ST-23 complex/Cluster A3 | | | |
| 868 | MDO1066 | | USA | 2001 | | Neisseria meningitidis | | 1626 | ST-269 complex | | | |
| 2281 | M7089 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | |
| 2299 | M7257 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | |
| 2316 | M7086 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 22 | ST-22 complex | 6 | 3 | |
| 2317 | M7084 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 22 | ST-22 complex | 6 | 3 | |
| 2322 | M7092 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1065 | ST-22 complex | 6 | 3 | |
| 2323 | M7100 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1286 | ST-22 complex | 6 | 3 | |
| 2324 | M7259 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1066 | ST-22 complex | 6 | 3 | |

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.

11.6.1 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.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Toggle: ☐ Field help: id

Search *Neisseria* PubMLST database

Isolate provenance/phenotype fields

Combine with: AND

country = USA +

year > 1999

Allele designations/scheme fields

Enter value... +

Action

Reset Submit

Display/sort options

Order by: id

Display: 25 records per page

Modify form options

Click to add or remove additional query terms:

- ☒ Hide Allele designations/scheme field values
- ☐ Show Allele designation status
- ☐ Show Tagged sequence status
- ☐ Show Filters

Save options

300 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page:

| Isolate fields | | | | | | | | MLST | | Finotyping antigens | | |
|----------------|------------|---------|---------|------|------------------------------|------------------------|-----------|------|-----------------------------|---------------------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 341 | M7085 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | | | |
| 499 | MDL01A0601 | | USA | 2001 | meningitis | Neisseria meningitidis | Y | 1378 | ST-23 complex/Cluster A3 | | | |
| 500 | MDL01A2447 | | USA | 2001 | invasive (unspecified/other) | Neisseria meningitidis | Y | 1379 | ST-23 complex/Cluster A3 | | | |
| 866 | MDO1227 | | USA | 2001 | | Neisseria meningitidis | | 1624 | ST-167 complex | | | |
| 867 | MDO1056 | | USA | 2001 | | Neisseria meningitidis | | 1625 | ST-23 complex/Cluster A3 | | | |
| 868 | MDO1066 | | USA | 2001 | | Neisseria meningitidis | | 1626 | ST-269 complex | | | |
| 2281 | M7089 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | |
| 2299 | M7257 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | |
| 2316 | M7086 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 22 | ST-22 complex | 6 | 3 | |
| 2317 | M7084 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 22 | ST-22 complex | 6 | 3 | |
| 2322 | M7092 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1065 | ST-22 complex | 6 | 3 | |
| 2323 | M7100 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1286 | ST-22 complex | 6 | 3 | |
| 2324 | M7259 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1066 | ST-22 complex | 6 | 3 | |

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.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Toggle: ☐ Field help: id

Search *Neisseria* PubMLST database

Isolate provenance/phenotype fields

Combine with: AND

country = USA +

year > 1999

Allele designations/scheme fields

ST (MLST) = 11 +

Action

Reset Submit

Display/sort options

Order by: id

Display: 25 records per page

Modify form options

Click to add or remove additional query terms:

- ☒ Hide Allele designations/scheme field values
- ☐ Show Allele designation status
- ☐ Show Tagged sequence status
- ☐ Show Filters

Save options

8 records returned. Click the hyperlinks for detailed information.

| Isolate fields | | | | | | | | MLST | | Finotyping antigens | | |
|----------------|---------|---------|---------|------|------------------------------|------------------------|-----------|------|-----------------------------|---------------------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 341 | M7085 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | | | |
| 2281 | M7089 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | |
| 2299 | M7257 | | USA | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | |
| 19369 | M13519 | | USA | 2005 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5-1 | 10-8 | F3-6 |
| 19371 | M15141 | | USA | 2006 | invasive (unspecified/other) | Neisseria meningitidis | C | 11 | ST-11 complex/ET-37 complex | 5-1 | 10-8 | |
| 19374 | M16917 | | USA | 2007 | invasive (unspecified/other) | Neisseria meningitidis | NG | 11 | ST-11 complex/ET-37 complex | 5-1 | 10-8 | |
| 19377 | M17661 | | USA | 2008 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5-1 | 10-8 | |
| 19379 | M18774 | | USA | 2009 | invasive (unspecified/other) | Neisseria meningitidis | W | 11 | ST-11 complex/ET-37 complex | 5 | 2 | F1-94 |

Analysis tools:

Breakdown:

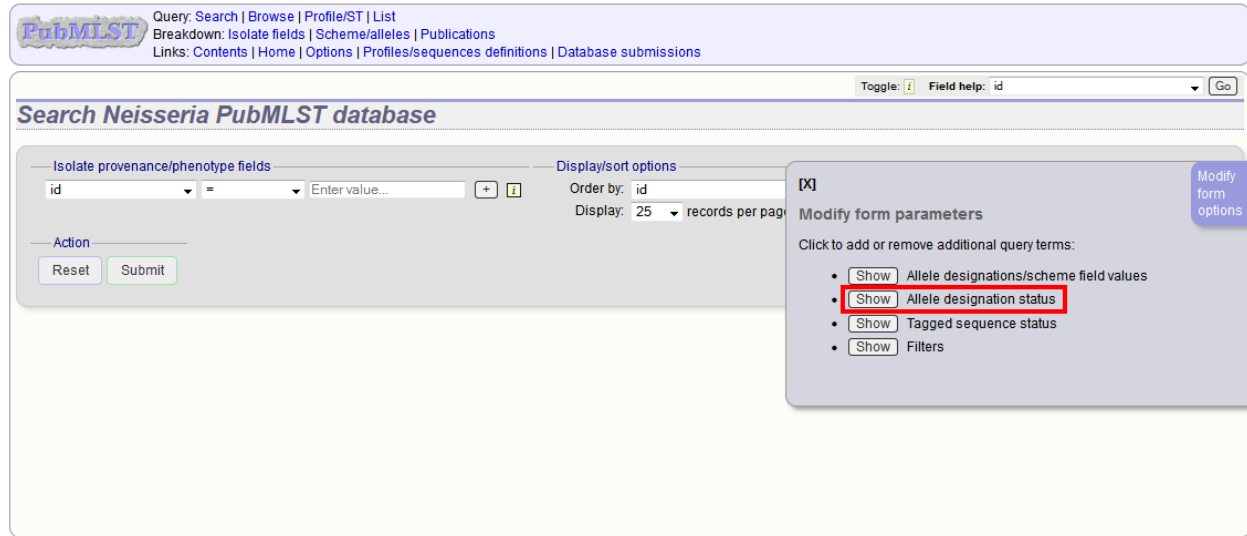
Analysis:

Export:

11.6.2 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 status fieldset is displayed by selecting it in the ‘Modify form options’ tab.



Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Search *Neisseria* PubMLST database

Isolate provenance/phenotype fields: id = Enter value... + ?

Display/sort options: Order by: id, Display: 25 records per page

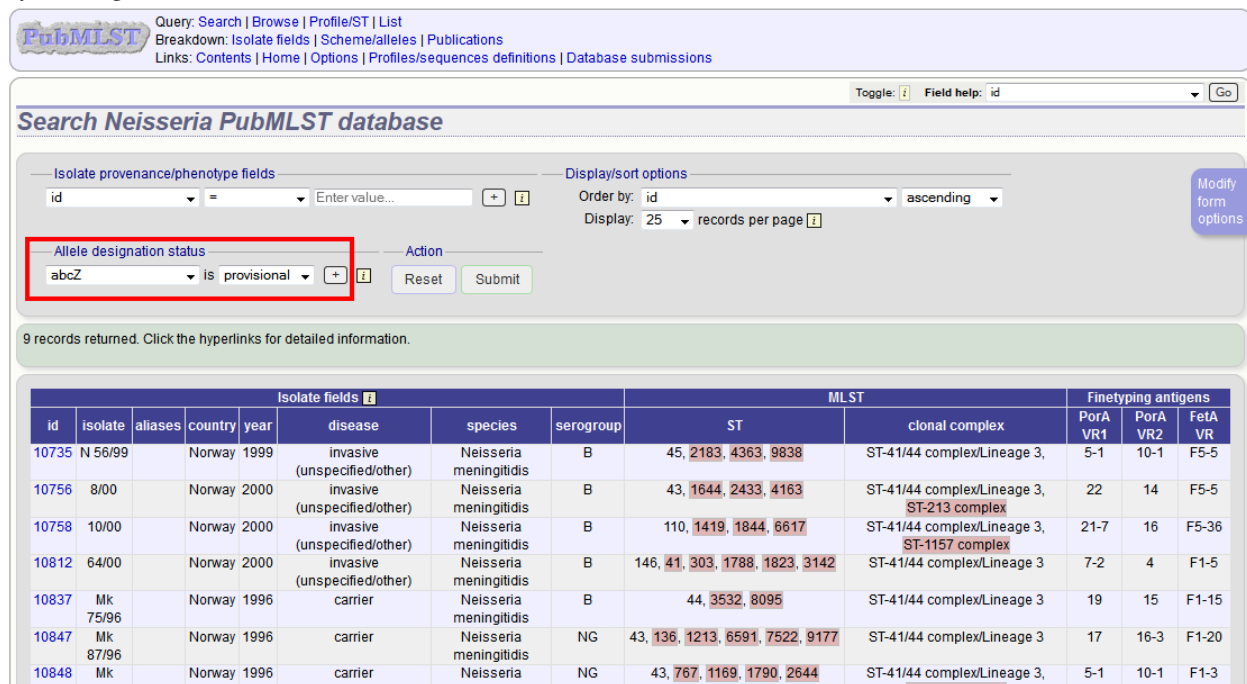
Action: Reset Submit

Modify form parameters

Click to add or remove additional query terms:

- Show Allele designations/scheme field values
- Show **Allele designation status**
- Show Tagged sequence status
- Show Filters

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



Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Search *Neisseria* PubMLST database

Isolate provenance/phenotype fields: id = Enter value... + ?

Display/sort options: Order by: id, ascending, Display: 25 records per page

Action: Reset Submit

Allele designation status: abcZ is provisional + ?

9 records returned. Click the hyperlinks for detailed information.

| Isolate fields | | | | | | | MLST | | Finotyping antigens | | | |
|----------------|----------|---------|---------|------|------------------------------|------------------------|-----------|---------------------------------|---|----------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 10735 | N 56/99 | | Norway | 1999 | invasive (unspecified/other) | Neisseria meningitidis | B | 45, 2183, 4363, 9838 | ST-41/44 complex/Lineage 3, | 5-1 | 10-1 | F5-5 |
| 10756 | 8/00 | | Norway | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 43, 1644, 2433, 4163 | ST-41/44 complex/Lineage 3, ST-213 complex | 22 | 14 | F5-5 |
| 10758 | 10/00 | | Norway | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 110, 1419, 1844, 6617 | ST-41/44 complex/Lineage 3, ST-1157 complex | 21-7 | 16 | F5-36 |
| 10812 | 64/00 | | Norway | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 146, 41, 303, 1788, 1823, 3142 | ST-41/44 complex/Lineage 3 | 7-2 | 4 | F1-5 |
| 10837 | Mk 75/96 | | Norway | 1996 | carrier | Neisseria meningitidis | B | 44, 3532, 8095 | ST-41/44 complex/Lineage 3 | 19 | 15 | F1-15 |
| 10847 | Mk 87/96 | | Norway | 1996 | carrier | Neisseria meningitidis | NG | 43, 136, 1213, 6591, 7522, 9177 | ST-41/44 complex/Lineage 3 | 17 | 16-3 | F1-20 |
| 10848 | Mk 88/96 | | Norway | 1996 | carrier | Neisseria meningitidis | NG | 43, 767, 1169, 1790, 2644 | ST-41/44 complex/Lineage 3, ST-427 complex | 5-1 | 10-1 | F1-3 |

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.

11.6.3 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.

The screenshot shows the 'Search Neisseria PubMLST database' interface. At the top, there are navigation links: Query: Search | Browse | Profile/ST | List; Breakdown: Isolate fields | Scheme/alleles | Publications; Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions. Below this is a search bar with 'Toggle: i Field help: id' and a 'Go' button. The main search area has a dropdown for 'Isolate provenance/phenotype fields' with 'id' selected, and a text input 'Enter value...'. To the right, 'Display/sort options' show 'Order by: id' and 'Display: 25 records per page'. A 'Modify form parameters' dialog is open on the right, with a 'Modify form options' button. The dialog lists query terms: 'Allele designations/scheme field values', 'Allele designation status', 'Tagged sequence status' (highlighted in red), and 'Filters'. Each term has a 'Show' button next to it.

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.

Sequence tag flags

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.

[Query](#) | [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Search *Neisseria* PubMLST database

Isolate provenance/phenotype fields

=

Display/sort options

Order by:
 Display: records per page

Action

[X] Modify form parameters

Click to add or remove additional query terms:

- Allele designations/scheme field values
- Allele designation status
- Tagged sequence status
- Filters**

- 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 may be complete or partial.
 - not started - The scheme has no loci with alleles designated.
- Sequence bin - Specify whether any sequence data has been associated with a record. Specific threshold values may be selected if these have been *set up for the database*.
 - Provenance fields - Dropdown list boxes of values for specific provenance fields may be present if set for the database. Users can choose to *add additional filters*.

11.7 Querying by allelic profile

If a scheme, such as MLST, has been defined for an isolate database it is possible to query the database against complete or partial allelic profiles. Even if no scheme is defined, queries can be made against all loci. This can also be done in sequence definition databases if the scheme has a primary key field defined.

On the index page, click ‘Search by combinations of loci (profiles)’ for any defined scheme. Enter either a partial (any combination of loci) or complete profile.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

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

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [?](#)

Search *Neisseria* PubMLST database by combinations of loci

Schemes
 Please select the scheme you would like to query:
 MLST

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:

Filters Project: ☐ Include old record versions

Options Search:

Display/sort options Order by: Display: records per page

Action

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

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [?](#)

Search *Neisseria* PubMLST database by combinations of loci

Schemes
 Please select the scheme you would like to query:
 MLST

Please enter your allelic profile below. Blank loci will be ignored. — Autofill profile by searching remote database —

| abcZ | adk | aroE | fumC | gdh | pdhC | pgm |
|------|-----|------|------|-----|------|-----|
| 2 | 3 | 4 | 3 | 8 | 4 | |

 ST:

Filters Project: ☐ Include old record versions

Options Search:

Display/sort options Order by: Display: records per page

Action

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

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Search *Neisseria* PubMLST database by combinations of loci

Schemes
Please select the scheme you would like to query:
MLST

Please enter your allelic profile below. Blank loci will be ignored. Autofill profile by searching remote database
 abcZ adk aroE fumC gdh pdhC pgm
 9 6 9 9 9 6 9 ST: 44

Filters
Project: ☐ Include old record versions

Options
Search:

Display/sort options
Order by:
 Display: 25

Action

Select the number of loci that you'd like to match in the options dropdown box. Available options are:

- Exact or nearest match
- Exact match only
- x or more matches
- y or more matches
- z or more matches

Where x,y, and z will range from n-1 to 1 where n is the number of loci in the scheme.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Search *Neisseria* PubMLST database by combinations of loci

Schemes
Please select the scheme you would like to query:
MLST

Please enter your allelic profile below. Blank loci will be ignored. Autofill profile by searching remote database
 abcZ adk aroE fumC gdh pdhC pgm
 9 6 9 9 9 6 9 ST: 44

Filters
Project: ☐ Include old record versions

Options
Search:

Display/sort options
Order by:
 Display: 25

Action

Click 'Submit'.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [?]

Search *Neisseria* PubMLST database by combinations of loci

Schemes

Please select the scheme you would like to query:

MLST

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 |
|------|-----|------|------|-----|------|-----|--|
| 9 | 6 | 9 | 9 | 9 | 6 | 9 | 44 <input type="button" value="Autofill"/> |

Filters — Options — Display/sort options

Project: Search: Order by:

☐ Include old record versions Display: records per page

Action

Exact matches found (7 loci).

119 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page:

| Isolate fields | | | | | | | MLST | | Finotyping antigens | | | |
|----------------------|---------|---------|-----------------|------|------------------------------|-------------------------------|-----------|----|----------------------------|----------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 41 | 19 | | Germany | 1999 | carrier | <i>Neisseria meningitidis</i> | B | 44 | ST-41/44 complex/Lineage 3 | | | |
| 70 | 38 | | Germany | 1999 | carrier | <i>Neisseria meningitidis</i> | NG | 44 | ST-41/44 complex/Lineage 3 | | | |
| 427 | NG E30 | Z4692 | Norway | 1988 | carrier | <i>Neisseria meningitidis</i> | B | 44 | ST-41/44 complex/Lineage 3 | 21 | 16 | F1-7 |
| 774 | 99 182 | | Canada | 1999 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | B | 44 | ST-41/44 complex/Lineage 3 | | | |
| 792 | 99-132 | | Canada | 1999 | invasive (unspecified/other) | <i>Neisseria meningitidis</i> | B | 44 | ST-41/44 complex/Lineage 3 | | | |
| 975 | 0069/93 | | Czech Republic | 1993 | carrier | <i>Neisseria meningitidis</i> | NG | 44 | ST-41/44 complex/Lineage 3 | 22 | 14-4 | F1-7 |
| 1097 | 3532 | Z7184 | The Netherlands | 1975 | | <i>Neisseria meningitidis</i> | B | 44 | ST-41/44 complex/Lineage 3 | | 4 | |
| 1603 | 0213/93 | | Czech Republic | 1993 | carrier | <i>Neisseria meningitidis</i> | B | 44 | ST-41/44 complex/Lineage 3 | 22 | 14-4 | F1-19 |
| 1604 | 0214/93 | | Czech Republic | 1993 | carrier | <i>Neisseria meningitidis</i> | NG | 44 | ST-41/44 complex/Lineage 3 | 22 | 14-4 | F1-7 |
| 1633 | 0244/93 | | Czech Republic | 1993 | carrier | <i>Neisseria meningitidis</i> | B | 44 | ST-41/44 complex/Lineage 3 | 22 | 14-4 | F1-7 |

11.8 Retrieving list of isolates or profiles

Both isolate and sequence definition databases can be queried against a list of values matching any criteria (isolate provenance fields, alleles, or scheme fields).

Click 'List query' on the main contents page.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

Select the attribute you wish to search against in the drop-down list box and enter the list of attributes in the box (one per line). Click ‘Submit’.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [?](#) Field help: [id](#) [Go](#)

Query Neisseria PubMLST database matching a field against a list

Please select attribute

isolate

Enter your list of attribute values below (one per line)

H44/76
MC58
alpha14

Filters

☐ Include old record versions

Display/sort options

Order by: [id](#) ascending

Display: 25 records per page

Action

[Reset](#) [Submit](#)

4 records returned. Click the hyperlinks for detailed information.

| Isolate fields | | | | | | | MLST | | Finotyping antigens | | | |
|----------------|---------|----------------|---------|------|------------------------------|------------------------|-----------|----|----------------------------|----------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 30 | 14 | alpha14 | Germany | 1999 | carrier | Neisseria meningitidis | NG | 53 | ST-53 complex | 7 | 30-3 | F5-5 |
| 237 | H44/76 | 44/76-3; Z3842 | Norway | 1976 | invasive (unspecified/other) | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | 7 | 16 | F3-3 |
| 240 | MC58 | Z7176 | UK | 1983 | | Neisseria meningitidis | B | 74 | ST-32 complex/ET-5 complex | 7 | 16-2 | F1-5 |
| 20477 | H44/76 | 44/76-3; Z3842 | Norway | 1976 | invasive (unspecified/other) | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | 7 | 16 | F3-3 |

Analysis tools:

Breakdown: [Fields](#) | [Two Field](#) | [Codons](#) | [Polymorphic sites](#) | [Combinations](#) | [Schemes/alleles](#) | [Publications](#) | [Sequence bin](#) | [Tag status](#)

Analysis: [BURST](#) | [Presence/Absence](#) | [Genome Comparator](#) | [BLAST](#)

Export: [Dataset](#) | [Contigs](#) | [Sequences](#)

11.9 Retrieving isolates by linked publication

Click ‘Publications’ in the Breakdown section of the contents page.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications**
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

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

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Publication breakdown of dataset

Filter query by: Author: Year: Display: Order by: records per page Action:

73 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page:

| PubMed id | Year | Citation | Title | Isolates in database |
|--------------------------|------|---|---|-------------------------------|
| 17517841 | 2007 | Taha MK, Vázquez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, Cafferkey MT, Carion F, Christensen JJ, Diggle M, Edwards G, Enriquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanoopoulos K, Kriz P, Lambertsen L, Levenet I, Musilek M, Paragi M, Sagner A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zarantonelli ML (2007) Antimicrob Agents Chemother 51: 2784-92 | Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis. | 1670 isolates |
| 18815379 | 2008 | Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Natl Acad Sci U S A 105: 15082-7 | Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis. | 1054 isolates |
| 15776372 | 2005 | Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M, Vogel U (2005) J Infect Dis 191: 1263-71 | Genetic analysis of meningococci carried by children and young adults. | 822 isolates |
| 15528708 | 2004 | Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Caugant DA, Maiden MC (2004) J Clin Microbiol 42: 5146-53 | Distribution of serogroups and genotypes among disease-associated and carried isolates of Neisseria meningitidis from the Czech Republic, Greece, and Norway. | 667 isolates |
| 17825091 | 2007 | Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007) BMC Biol 5: 35 | Species status of Neisseria gonorrhoeae: evolutionary and epidemiological inferences from multilocus sequence typing. | 576 isolates |
| 15537808 | 2005 | Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005) Mol Biol Evol 22: 562-9 | The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in Neisseria meningitidis. | 378 isolates |
| 18375809 | 2008 | Russell JE, Urwin R, Grav SJ, Fox AJ, Feavers IM, Maiden MC (2008) Microbiol Rev 72: 154-64 | Molecular epidemiology of meningococcal disease in England and Wales. | 323 isolates |

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

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Publication breakdown of dataset

Filter query by
 Author:
 Year:

Display
 Order by:
 Display: records per page

Action

13 records returned. Click the hyperlinks for detailed information.

| PubMed id | Year | Citation | Title | Isolates in database |
|--------------------------|------|--|--|-------------------------------|
| 17517841 | 2007 | Taha MK, Vázquez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, Cafferkey MT, Carion F, Christensen JJ, Diggle M, Edwards G, Enriquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lamberts L, Levenet I, Musilek M, Paragi M, Sager A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zaranonelli ML (2007) Antimicrob Agents Chemother 51: 2784-92 | Target gene sequencing to characterize the penicillin G susceptibility of <i>Neisseria meningitidis</i> . | 1670 isolates |
| 18815379 | 2008 | Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Natl Acad Sci U S A 105: 15082-7 | Role of selection in the emergence of lineages and the evolution of virulence in <i>Neisseria meningitidis</i> . | 1054 isolates |
| 15776372 | 2005 | Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M, Vogel U (2005) J Infect Dis 191: 1263-71 | Genetic analysis of meningococci carried by children and young adults. | 822 isolates |
| 15528708 | 2004 | Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Caugant DA, Maiden MC (2004) J Clin Microbiol 42: 5146-53 | Distribution of serogroups and genotypes among disease-associated and carried isolates of <i>Neisseria meningitidis</i> from the Czech Republic, Greece, and Norway. | 667 isolates |
| 17825091 | 2007 | Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007) BMC Biol 5: 35 | Species status of <i>Neisseria gonorrhoeae</i> : evolutionary and epidemiological inferences from multilocus sequence typing. | 576 isolates |
| 15537808 | 2005 | Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005) Mol Biol Evol 22: 562-9 | The influence of mutation, recombination, | 378 isolates |

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

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Publication breakdown of dataset

Filter query by
 Author:
 Year:

Display
 Order by:
 Display: records per page

Action

13 records returned. Click the hyperlinks for detailed information.

| PubMed id | Year | Citation | Title | Isolates in database |
|--------------------------|------|--|--|-------------------------------|
| 17517841 | 2007 | Taha MK, Vázquez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, Cafferkey MT, Carion F, Christensen JJ, Diggle M, Edwards G, Enriquez R, Fazio C, Frosch M, Heuberger S, Hoffmann S, Jolley KA, Kadlubowski M, Kechrid A, Kesanopoulos K, Kriz P, Lamberts L, Levenet I, Musilek M, Paragi M, Sager A, Skoczynska A, Stefanelli P, Thulin S, Tzanakaki G, Unemo M, Vogel U, Zaranonelli ML (2007) Antimicrob Agents Chemother 51: 2784-92 | Target gene sequencing to characterize the penicillin G susceptibility of <i>Neisseria meningitidis</i> . | 1670 isolates |
| 18815379 | 2008 | Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC (2008) Proc Natl Acad Sci U S A 105: 15082-7 | Role of selection in the emergence of lineages and the evolution of virulence in <i>Neisseria meningitidis</i> . | 1054 isolates |
| 15776372 | 2005 | Claus H, Maiden MC, Wilson DJ, McCarthy ND, Jolley KA, Urwin R, Hessler F, Frosch M, Vogel U (2005) J Infect Dis 191: 1263-71 | Genetic analysis of meningococci carried by children and young adults. | 822 isolates |
| 15528708 | 2004 | Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Caugant DA, Maiden MC (2004) J Clin Microbiol 42: 5146-53 | Distribution of serogroups and genotypes among disease-associated and carried isolates of <i>Neisseria meningitidis</i> from the Czech Republic, Greece, and Norway. | 667 isolates |
| 17825091 | 2007 | Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007) BMC Biol 5: 35 | Species status of <i>Neisseria gonorrhoeae</i> : evolutionary and epidemiological inferences from multilocus sequence typing. | 576 isolates |
| 15537808 | 2005 | Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005) Mol Biol Evol 22: 562-9 | The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in <i>Neisseria meningitidis</i> . | 378 isolates |
| 15784588 | 2005 | Bennett JS, Griffiths DT, McCarthy ND, Sleeman KL, Jolley KA, Crook DW, Maiden MC (2005) Infect Immun 73: 2424-32 | Genetic diversity and carriage dynamics of <i>Neisseria lactamica</i> in infants. | 271 isolates |
| 11101585 | 2000 | Jolley KA, Kalmusova J, Feil FJ, Gupta S, Musilek M, Kriz P, Maiden MC (2000) J Clin Microbiol 38: 4492-8 | Carried meningococci in the Czech Republic: a | 217 isolates |

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



Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Publications cited in the *Neisseria* PubMLST database

Citation query (PubMed id: 17825091)


Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007) *BMC Biol* 5:35

Species status of *Neisseria gonorrhoeae*: evolutionary and epidemiological inferences from multilocus sequence typing.

The seven loci *Neisseria* MLST scheme was readily adapted to *N. gonorrhoeae* isolates, providing a highly discriminatory typing method. In addition, these data permitted phylogenetic and population genetic inferences to be made, including direct comparisons with *N. meningitidis* and *N. lactamica*. Examination of these data demonstrated that alleles were rarely shared among the three species. Analysis of variation at a single locus, *gdh*, provided a rapid means of identifying misclassified isolates and determining whether mixed cultures were present.

576 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [>](#) [Last](#)

| Isolate fields  | | | | | | | | ML ST | | Finotyping antigens | | |
|--|----------|-------------|--------------|------|------------------------------|------------------------|-----------|-------|---------------------------|---------------------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 1 | A4/M1027 | B1; Z1001 | USA | 1937 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 5-2 | 10 | F1-5 |
| 2 | 120M | B35; Z1035 | Pakistan | 1967 | meningitis and septicaemia | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F5-1 |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F3-6 |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F5-1 |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 31 | 10 | B269; Z1269 | Burkina Faso | 1963 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 34 | 20 | B275; Z1275 | Niger | 1963 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F1-7 |
| 35 | 26 | B278; Z1278 | Niger | 1963 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13 | F1-5 |
| 46 | 255 | B318; Z1318 | Burkina Faso | 1966 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7-2 | 13-1 | F1-5 |
| 52 | 243 | B362; Z1362 | Cameroon | 1966 | | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13 | F1-5 |
| 61 | 393 | B392; Z1392 | Greece | 1968 | carrier | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F5-1 |
| 64 | 254 | B439; Z1439 | Djibouti | 1966 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F1-7 |
| 67 | S5611 | B466; Z1466 | Australia | 1977 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F5-1 |
| 82 | 11-004 | B503; Z1503 | China | 1984 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-8 |

11.10 User-configurable options

The BIGSdb user interface is configurable in a number of ways. Choices made are remembered between sessions. If the database requires you to log on, the options are associated with your user account, whereas if it is a public database, that you haven't logged in to, the options are associated with a browser cookie so they will be remembered if you connect from the same computer (using the same browser).

Most options are set by clicking the 'Set general options' link on the database contents page. Most of the available options are visible for isolate databases, whereas sequence definition databases have fewer available.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options** including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- [Update history](#)
- [About BIGSdb](#)

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

11.10.1 General options

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

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Set database options

Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

General options

Display **25** records per page.
Page bar position: top and bottom
Display **100** nucleotides per line in sequence alignments.
Display **100** nucleotides of flanking sequence (where available).
☒ Display locus aliases if set.
☒ Enable tooltips (beginner's mode).

Main results table

Isolate record display

Provenance field display

Query filters

Reset

Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

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 than the locus sequence.
- Locus aliases - Loci can have multiple names (aliases). Setting this option will display all alternative names in results tables.
- Tooltips (beginner's mode) - Most query forms have help available in the form of information tooltips. These can be switched on/off here. They can also be toggled off by clicking the Toggle: 'i' button at the top-right of the display of some pages.

Click 'Set options' to remember any changes you make.

11.10.2 Main results table

The 'main results table' tab contains options for the display of paged results following a query.

Click the 'Main results table' header to display the tab.


[Query: Search | Browse | Profile/ST | List](#)
[Breakdown: Isolate fields | Scheme/alleles | Publications](#)
[Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions](#)

Set database options

Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

▼ General options

Display records per page.
Page bar position:
Display nucleotides per line in sequence alignments.
Display nucleotides of flanking sequence (where available).
☒ Display locus aliases if set.
☒ Enable tooltips (beginner's mode).

► Main results table

► Isolate record display

► Provenance field display

► Query filters

Reset

Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

The 'main results table' tab will scroll up.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Set database options

Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

General options

Main results table

- ☐ Hyperlink allele designations where possible.
- ☒ Differentiate provisional allele designations.
- ☐ Display information about sequence bin records tagged with locus information (tooltip).
- ☐ Display sequence bin size.
- ☐ Display contig count.
- ☐ Display publications.

Set options

Isolate record display

Provenance field display

Query filters

Reset

Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

Reset all to defaults

This tab allows the following options to be modified:

- Hyperlink allele designations - Hyperlinks point to an information page about the particular allele definition. Depending on the locus, these may exist on a different website.
- Differentiate provisional allele designations - Allele designations can be set as confirmed or provisional, usually depending on the method of assignment. Selecting this option will display provisional designations in a different colour to confirmed designations.
- Information about sequence bin records - Creates a tooltip that displays details about sequence tags corresponding to a locus.
- Sequence bin records - Displays a tooltip linking to the sequence tag if available.
- Sequence bin size - Displays the size of the sum of all contigs associated with each isolate record.
- Contig count - Displays the number of contigs associated with each isolate record.
- Publications - Displays citations with links to PubMed for each record.

11.10.3 Isolate record display

The 'isolate record display' tab contains options for the display of a full isolate record.

Click the 'Isolate record display' tab to display the tab.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Set database options

Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

▼ General options

Display records per page.
 Page bar position:
 Display nucleotides per line in sequence alignments.
 Display nucleotides of flanking sequence (where available).
☒ Display locus aliases if set.
☒ Enable tooltips (beginner's mode).

► Main results table

► Isolate record display

► Provenance field display

► Query filters

Reset
 Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

The 'Isolate record display' tab will scroll up.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Set database options

Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

► General options

► Main results table

▼ Isolate record display

☒ Differentiate provisional allele designations.
☐ Display sender, curator and last updated details for allele designations (tooltip).
☒ Display information about sequence bin records tagged with locus information (tooltip).
☐ Display information about whether alleles have flags defined in sequence definition database (shown in sequence detail tooltip).
☒ Display full information about sample records (tooltip).

► Provenance field display

► Query filters

Reset
 Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

This tab allows the following options to be modified:

- Differentiate provisional allele designations - Allele designations can be set as confirmed or provisional, usually depending on the method of assignment. Selecting this option will display provisional designations in a different colour to confirmed designations.

- Display sender, curator and last updated records - Displays a tooltip containing sender information next to each allele designation.
- Sequence bin information - Displays a tooltip with information about the position of the sequence if tagged within the sequence bin.
- Allele flags - Displays information about whether alleles have flags defined in sequence definition databases.
- Display full information about sample records - Used when the database is used as part of a basic laboratory information management system (LIMS). This option will display records of samples available for the displayed isolate.

11.10.4 Provenance field display

The ‘provenance field display’ tab contains checkboxes for fields to display in the main results table.

Click the ‘Provenance field display’ tab to display the tab.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Set database options

Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

▼ General options

Display records per page.
 Page bar position:
 Display nucleotides per line in sequence alignments.
 Display nucleotides of flanking sequence (where available).
☒ Display locus aliases if set.
☒ Enable tooltips (beginner's mode).

► Main results table
 ► Isolate record display
 ► **Provenance field display**
 ► Query filters

Reset
 Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

The ‘Provenance field display’ tab will scroll up.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Set database options

Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

General options

Main results table

Isolate record display

Provenance field display

Select the isolate provenance fields that you wish to be displayed in the main results table following a query. Settings for displaying locus and scheme data can be made by performing a locus, scheme or scheme field query and then selecting the 'Customize' option.

| | | | |
|---|---|--|--|
| <input checked="" type="checkbox"/> isolate | <input checked="" type="checkbox"/> disease | <input type="checkbox"/> amoxicillin | <input type="checkbox"/> ciprofloxacin_range |
| <input checked="" type="checkbox"/> aliases | <input type="checkbox"/> source | <input type="checkbox"/> sulphonamide | <input type="checkbox"/> pending_assembly |
| <input type="checkbox"/> strain_designation | <input type="checkbox"/> epidemiology | <input type="checkbox"/> ceftriaxone | <input type="checkbox"/> assembly_status |
| <input checked="" type="checkbox"/> country | <input checked="" type="checkbox"/> species | <input type="checkbox"/> ceftriaxone_range | <input type="checkbox"/> ENA_accession |
| <input type="checkbox"/> continent | <input checked="" type="checkbox"/> serogroup | <input type="checkbox"/> chloramphenicol | <input type="checkbox"/> private_project |
| <input type="checkbox"/> region | <input type="checkbox"/> MLEE_designation | <input type="checkbox"/> chloramphenicol_range | <input type="checkbox"/> comments |
| <input checked="" type="checkbox"/> year | <input type="checkbox"/> serotype | <input type="checkbox"/> cefotaxime | <input type="checkbox"/> sender |
| <input type="checkbox"/> epidemiological_year | <input type="checkbox"/> sero_subtype | <input type="checkbox"/> cefotaxime_range | <input type="checkbox"/> curator |
| <input type="checkbox"/> age_yr | <input type="checkbox"/> ET_no | <input type="checkbox"/> rifampicin | <input type="checkbox"/> date_entered |
| <input type="checkbox"/> age_mth | <input type="checkbox"/> penicillin | <input type="checkbox"/> rifampicin_range | <input type="checkbox"/> timestamp |
| <input type="checkbox"/> sex | <input type="checkbox"/> penicillin_range | <input type="checkbox"/> ciprofloxacin | |

Query filters

Reset

Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

Some fields will be checked by default - these are defined during *database setup* (maindisplay option).

Check any fields that you wish to be displayed and then click 'Set options'. You can return to the default selection by clicking 'Default' followed by 'Set options'.

11.10.5 Query filters

The 'query filters' tab contains checkboxes for provenance fields and scheme completion status. Checking these results in drop-down list box filters appearing in the query page *filters fieldset*.

Click the 'Query filters' tab to display the tab.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Set database options

Here you can set options for your use of the website. Options are remembered between sessions and affect the current database (Neisseria PubMLST) only. If some of the options don't appear to set when you next go to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

▼ General options

Display records per page.
 Page bar position:
 Display nucleotides per line in sequence alignments.
 Display nucleotides of flanking sequence (where available).
☒ Display locus aliases if set.
☒ Enable tooltips (beginner's mode).

► Main results table

► Isolate record display

► Provenance field display

► **Query filters**

Reset
 Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

The 'Query filters' tab will scroll up.

► Main results table

► Isolate record display

► Provenance field display

▼ **Query filters**

Select the fields for which you would like dropdown lists containing known values on which to filter query results. These will be available in the filters section of the query interface.

| | | |
|--|--|--|
| <input type="checkbox"/> isolate | <input type="checkbox"/> ENA_accession | <input type="checkbox"/> LOS alpha chain transferases profile completion |
| <input type="checkbox"/> country | <input type="checkbox"/> private_project | <input type="checkbox"/> LOS inner core transferases profile completion |
| <input type="checkbox"/> continent | <input type="checkbox"/> comments | <input type="checkbox"/> LOS transport/export profile completion |
| <input type="checkbox"/> region | <input type="checkbox"/> sender | <input type="checkbox"/> Lipid A biosynthesis: acyltransferases profile completion |
| <input type="checkbox"/> year | <input type="checkbox"/> curator | <input type="checkbox"/> Lipid A biosynthesis: other profile completion |
| <input type="checkbox"/> epidemiological_year | <input type="checkbox"/> date_entered | <input type="checkbox"/> N. meningitidis profile completion |
| <input type="checkbox"/> age_yr | <input type="checkbox"/> datestamp | <input type="checkbox"/> Neisseria genus core genes profile completion |
| <input type="checkbox"/> age_mth | <input checked="" type="checkbox"/> MLST profile completion | <input type="checkbox"/> Nucleotide excision repair profile completion |
| <input type="checkbox"/> sex | <input type="checkbox"/> Finetyping antigens profile completion | <input type="checkbox"/> Pilus genes profile completion |
| <input type="checkbox"/> disease | <input type="checkbox"/> 16S profile completion | <input type="checkbox"/> Protein glycosylation profile completion |
| <input type="checkbox"/> source | <input type="checkbox"/> ADP-heptose biosynthesis profile completion | <input type="checkbox"/> Purine metabolism profile completion |
| <input type="checkbox"/> epidemiology | <input type="checkbox"/> Aminoacyl-tRNA biosynthesis profile completion | <input type="checkbox"/> Pyrimidine metabolism profile completion |
| <input type="checkbox"/> species | <input type="checkbox"/> Antibiotic resistance profile completion | <input type="checkbox"/> Pyruvate dehydrogenase complex profile completion |
| <input type="checkbox"/> serogroup | <input type="checkbox"/> Antigen genes profile completion | <input type="checkbox"/> RNA polymerase profile completion |
| <input type="checkbox"/> MLEE_designation | <input type="checkbox"/> Capsule Region A - Serogroup A profile completion | <input checked="" type="checkbox"/> Ribosomal MLST profile completion |
| <input type="checkbox"/> serotype | <input type="checkbox"/> Capsule Region A - Serogroup B profile completion | <input type="checkbox"/> Small regulatory RNAs profile completion |
| <input type="checkbox"/> sero_subtype | <input type="checkbox"/> Capsule Region A - Serogroup C profile completion | <input type="checkbox"/> TCA cycle profile completion |
| <input type="checkbox"/> ET_no | <input type="checkbox"/> Capsule Region A - Serogroup E profile completion | <input type="checkbox"/> UDP-GlcNAc synthesis profile completion |
| <input type="checkbox"/> penicillin | <input type="checkbox"/> Capsule Region A - Serogroup H profile completion | <input type="checkbox"/> UDP-glucose and UDP-galactose biosynthesis profile completion |
| <input type="checkbox"/> penicillin_range | <input type="checkbox"/> Capsule Region A - Serogroup L profile completion | <input type="checkbox"/> completion |
| <input type="checkbox"/> amoxicillin | <input type="checkbox"/> Capsule Region A - Serogroup W profile completion | <input type="checkbox"/> VirB T4SS profile completion |
| <input type="checkbox"/> sulphonamide | <input type="checkbox"/> Capsule Region A - Serogroup X profile completion | <input type="checkbox"/> beta lactamase plasmid profile completion |
| <input type="checkbox"/> ceftriaxone | <input type="checkbox"/> Capsule Region A - Serogroup Y profile completion | <input type="checkbox"/> eMLST (20 locus partial genes) profile completion |
| <input type="checkbox"/> ceftriaxone_range | <input type="checkbox"/> Capsule Region A - Serogroup Z profile completion | <input type="checkbox"/> eMLST (20 locus whole genes) profile completion |
| <input type="checkbox"/> chloramphenicol | <input type="checkbox"/> Capsule Region B profile completion | <input type="checkbox"/> rplF species profile completion |
| <input type="checkbox"/> chloramphenicol_range | <input type="checkbox"/> Capsule Region C profile completion | <input checked="" type="checkbox"/> Publications |
| <input type="checkbox"/> cefotaxime | <input type="checkbox"/> Capsule Region D and D' profile completion | |
| <input type="checkbox"/> cefotaxime_range | <input type="checkbox"/> Conjugative Plasmid profile completion | |
| <input type="checkbox"/> rifampicin | <input type="checkbox"/> DNA replication profile completion | |
| <input type="checkbox"/> rifampicin_range | <input type="checkbox"/> Factor H-binding protein profile completion | |
| <input type="checkbox"/> ciprofloxacin | <input type="checkbox"/> Glycolysis profile completion | |
| <input type="checkbox"/> ciprofloxacin_range | <input type="checkbox"/> Gonococcal Genetic Island profile completion | |
| <input type="checkbox"/> pending_assembly | <input type="checkbox"/> Iron acquisition profile completion | |
| <input type="checkbox"/> assembly_status | <input type="checkbox"/> Kdo addition profile completion | |

A list of possible filters appears. Click any checkbox for a filter you would like to make available. Click ‘Set options’ when done. You can return to the default selection by clicking ‘Default’ followed by ‘Set options’.

11.10.6 Modifying locus and scheme display options

Whether or not loci, schemes or scheme fields are displayed in result tables, isolate records, or within query dropdown boxes can all be set with default options when first defined. These attributes can, however, be overridden by a user, and these selections will be remembered between sessions.

The procedure to modify these attributes is the same for locus, schemes or scheme fields, so the steps for loci will be demonstrated only.

Click the appropriate link on the isolate contents page.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- [Search database](#) - advanced queries.
- [Browse database](#) - peruse all records.
- [Search by combinations of loci \(profiles\)](#) - including partial matching.
- [List query](#) - find isolates by matching a field to an entered list
- [Projects](#) - main projects defined in database.

Option settings

- [Set general options](#) - including [isolate table field handling](#).
- Set display and query options for **locus, schemes or scheme fields**.

Submissions

- [Manage submissions](#)

General information

- Isolates: 34221
- Last updated: 2015-07-02
- [Update history](#)
- [About BIGSdb](#)

Breakdown

- [Single field](#)
- [Two field](#)
- [Unique combinations](#)
- [Scheme and alleles](#)
- [Publications](#)
- [Sequence bin](#)

Export

- [Export dataset](#)
- [Contigs](#)
- [Sequences](#) - XMFA / concatenated FASTA formats

Analysis

- [Codon usage](#)
- [Presence/absence status of loci](#)
- [Genome comparator](#)
- [BLAST](#)

Miscellaneous

- [Description of database fields](#)

Either select the locus id by querying for it directly.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Query loci for Neisseria PubMLST database

Please enter your search criteria below (or leave blank and submit to return all records). Matching loci will be returned and you will then be able to update their display and query settings.

Search criteria

id
=
abcZ

+

?

Display
Order by: id ascending
Display: 25 records per page

Filter query by:

Action
Reset Submit

1 record returned. Click the hyperlink for detailed information.

Customize
locus options

| id | data type | allele id format | length | length varies | coding sequence | orf | genome position | isolate display* | main display* | query field* | analysis* |
|------|-----------|------------------|--------|---------------|-----------------|-----|-----------------|------------------|---------------|--------------|-----------|
| abcZ | DNA | integer | 433 | false | true | 1 | 1176340 | allele only | false | true | true |

* Default values are displayed for this field. These may be overridden by user preference.

Designations can be queried using *standard operators*.

Alternatively, you can search by filtering loci by schemes. Click the ‘Filter query by’ header and select the scheme in the dropdown box.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [?](#)

Query loci for *Neisseria* PubMLST database

Please enter your search criteria below (or leave blank and submit to return all records). Matching loci will be returned and you will then be able to update their display and query settings.

Search criteria: id = [] + [?](#)

Display: Order by: id ascending | Display: 25 records per page [?](#)

Filter query by

data type: []
 allele id format: []
 length varies: [] [?](#)
 coding sequence: []
 orf: []
 match longest: [] [?](#)
 pcr filter: [] [?](#)
 probe filter: [] [?](#)
 flag table: [] [?](#)
 isolate display: []
 main display: [] [?](#)
 query field: [] [?](#)
 analysis: [] [?](#)
 curator: [] [?](#)
 scheme: MLST [?](#)

Action: [Reset](#) [Submit](#)

7 records returned. Click the hyperlinks for detailed information.

Customize: [locus options](#)

| id | data type | allele id format | length | length varies | coding sequence | orf | genome position | isolate display* | main display* | query field* | analysis* |
|----------------------|-----------|------------------|--------|---------------|-----------------|-----|-----------------|------------------|---------------|--------------|-----------|
| abcZ | DNA | integer | 433 | false | true | 1 | 1176340 | allele only | false | true | true |
| adk | DNA | integer | 465 | false | true | 1 | 991951 | allele only | false | true | true |
| aroE | DNA | integer | 490 | false | true | 2 | 2079469 | allele only | false | true | true |
| fumC | DNA | integer | 465 | false | true | 1 | 1592943 | allele only | false | true | true |
| gdh | DNA | integer | 501 | false | true | 1 | 1514394 | allele only | false | true | true |
| pdhC | DNA | integer | 480 | false | true | 1 | 1453970 | allele only | false | true | true |
| pgm | DNA | integer | 450 | false | true | 1 | 965481 | allele only | false | true | true |

* Default values are displayed for this field. These may be overridden by user preference.

Once loci have been selected, click Customize ‘locus options’.

7 records returned. Click the hyperlinks for detailed information.

Customize: [locus options](#)

| id | data type | allele id format | length | length varies | coding sequence | orf | genome position | isolate display* | main display* | query field* | analysis* |
|----------------------|-----------|------------------|--------|---------------|-----------------|-----|-----------------|------------------|---------------|--------------|-----------|
| abcZ | DNA | integer | 433 | false | true | 1 | 1176340 | allele only | false | true | true |
| adk | DNA | integer | 465 | false | true | 1 | 991951 | allele only | false | true | true |
| aroE | DNA | integer | 490 | false | true | 2 | 2079469 | allele only | false | true | true |
| fumC | DNA | integer | 465 | false | true | 1 | 1592943 | allele only | false | true | true |
| gdh | DNA | integer | 501 | false | true | 1 | 1514394 | allele only | false | true | true |
| pdhC | DNA | integer | 480 | false | true | 1 | 1453970 | allele only | false | true | true |
| pgm | DNA | integer | 450 | false | true | 1 | 965481 | allele only | false | true | true |

* Default values are displayed for this field. These may be overridden by user preference.

You can then choose to add or remove individual loci from the selection by clicking the appropriate checkboxes. At the bottom of the page are a number of attributes that you can change - clicking ‘Change’ will affect all selected loci.

Possible options for loci are:

- isolate_display - Sets how the locus is displayed within an isolate record:

- allele only - display only identifier
 - sequence - display the full sequence
 - hide - don't show at all
- main_display - Sets whether the locus is displayed in the main results table following a query.
- query_field - Sets whether the locus appears in dropdown list boxes to be used within queries.
- analysis - Sets whether the locus can be used in data analysis functions.

Note: Settings for loci can be overridden by those set for schemes that they are members of. For example, if you set a locus to be displayed within a main results table, but that locus is a member of a scheme and you set that scheme not to be displayed, then the locus will not be shown. Conversely, if you set a scheme to be displayed, but set its member locus not to be shown, then that locus will not be displayed (but other loci and scheme fields may be, depending on their independent settings).

Data analysis plugins

12.1 Locus explorer

The locus explorer is a sequence definition database plugin. It can create schematics showing the polymorphic sites within a locus, calculate the GC content and generate aligned translated sequences.

Click 'Locus Explorer' from the sequence definition database contents page.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Toggle: ⓘ

Neisseria locus/sequence definitions database

The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.

Query database

- Sequence query - query an allele sequence.
- Batch sequence query - query multiple sequences in FASTA format.
- Sequence attribute search - find alleles by matching attributes.
- Browse profiles
- Search profiles
- List - find profiles matched to entered list.
- Search by combinations of alleles - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

Downloads

- Allele sequences
- MLST

Option settings

- Set general options

Submissions

- Manage submissions

General information

- Number of sequences: 519870
- Number of profiles: [Show](#)
- Last updated: 2015-07-02
- [Profile update history](#)
- [About BIGSdb](#)

Export

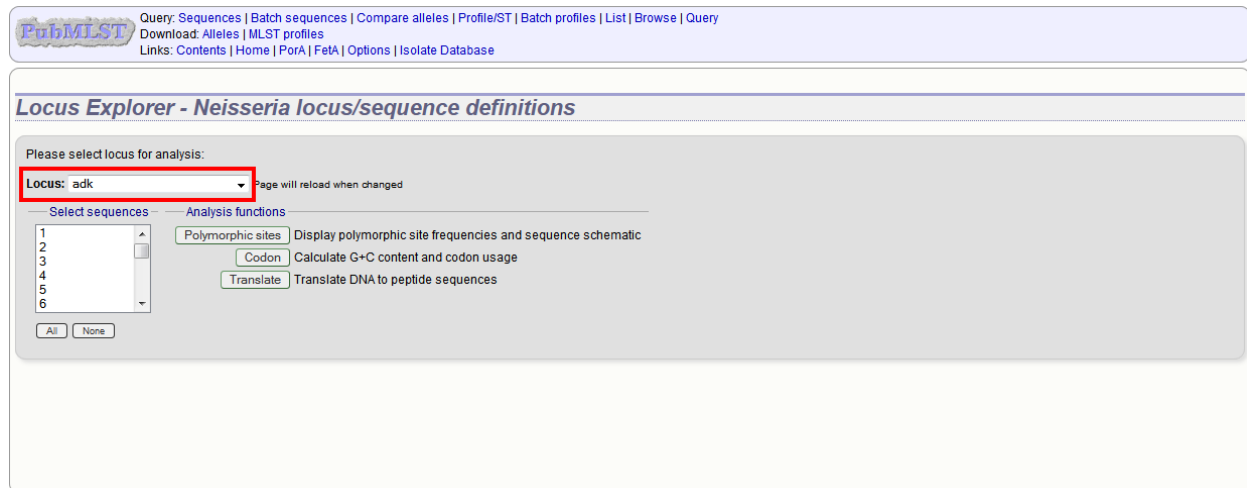
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Sequence similarity - find sequences most similar to selected allele.
- Sequence comparison - display a comparison between two sequences.
- Locus Explorer** - tool for analysing allele sequences stored for particular locus.

12.1.1 Polymorphic site analysis

Select the locus you would like to analyse in the Locus dropdown box. The page will reload.



PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Locus Explorer - *Neisseria* locus/sequence definitions

Please select locus for analysis:
Locus: adk Page will reload when changed

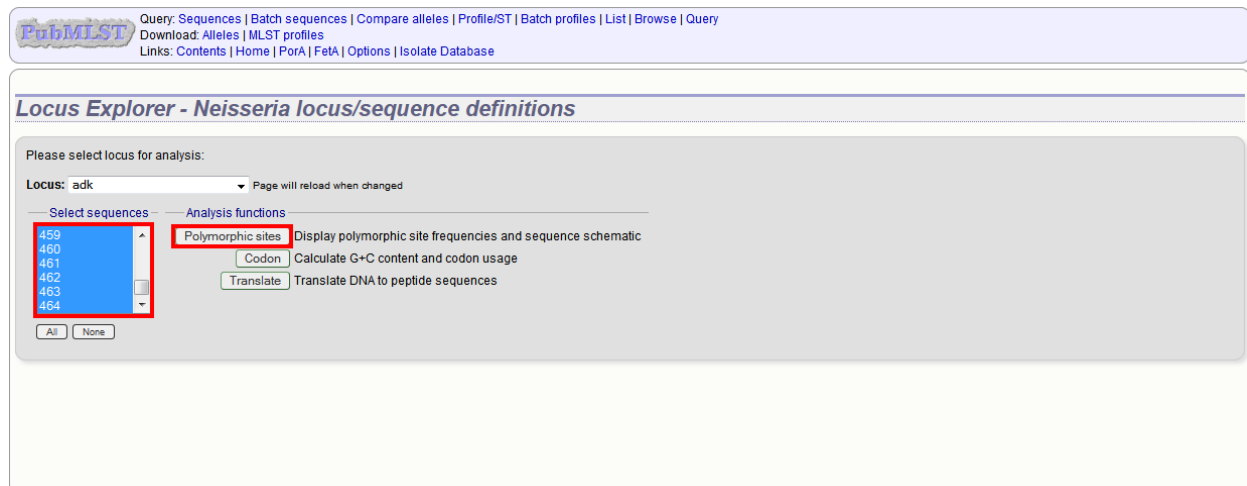
Select sequences Analysis functions

1
2
3
4
5
6

Polymorphic sites Display polymorphic site frequencies and sequence schematic
Codon Calculate G+C content and codon usage
Translate Translate DNA to peptide sequences

All None

Select the alleles that you would like to include in the analysis. Variable length loci are limited to 200 sequences or fewer since these need to be aligned. Click 'Polymorphic sites'.



PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Locus Explorer - *Neisseria* locus/sequence definitions

Please select locus for analysis:
Locus: adk Page will reload when changed

Select sequences Analysis functions

459
460
461
462
463
464

Polymorphic sites Display polymorphic site frequencies and sequence schematic
Codon Calculate G+C content and codon usage
Translate Translate DNA to peptide sequences

All None

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.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Polymorphic site analysis

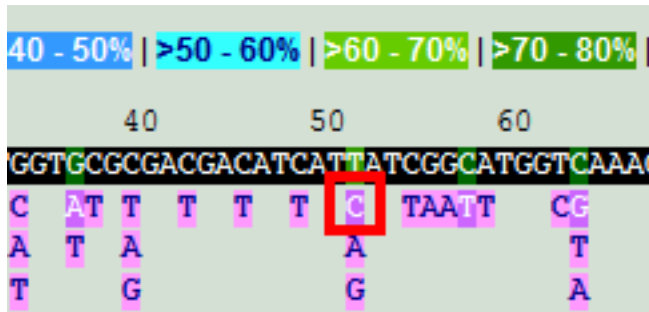
adk

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 information. The width of the display can be altered by going to the options page - change this if the display goes off the page.

464 alleles included in analysis. 265 polymorphic sites found.

Key: 0 - 10% | >10 - 20% | >20 - 30% | >30 - 40% | >40 - 50% | >50 - 60% | >60 - 70% | >70 - 80% | >80 - 90% | >90 - 100%

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



PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Site Explorer

adk position 51

464 alleles included in analysis.

| Base | Number of alleles | Percentage of alleles | MLST profiles |
|------|-------------------|-----------------------|------------------------|
| T | 370 | 79.74 | 10619 / 10883 (97.57%) |
| C | 92 | 19.83 | 262 / 10883 (2.41%) |
| A | 1 (adk-351) | 0.22 | 1 / 10883 (0.01%) |
| G | 1 (adk-413) | 0.22 | 1 / 10883 (0.01%) |

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

401 410 420 430 440 450 460

TTTACGCAACTGGAAGGCGACACGGCGCTAATACATCAAGTTGACGGCACACGCGTA

AC GTC TCG TACACCA TAGGTCT TAT CG T T GA T T G GATTC G

C G G CACA GG CAG C C G A A G TTAC T

T A A A T C

Nucleotide frequencies

| Position | Nucleotide | | | | | | | | | |
|----------|------------|-----|-----|-----|-------|-------|-------|-------|---|---|
| | A | C | G | T | %A | %C | %G | %T | % | % |
| 2 | 463 | 0 | 1 | 0 | 99.78 | 0.22 | | | | |
| 3 | 462 | 0 | 2 | 0 | 99.57 | 0.43 | | | | |
| 4 | 1 | 0 | 463 | 0 | 0.22 | 99.78 | | | | |
| 5 | 0 | 463 | 1 | 0 | | 99.78 | 0.22 | | | |
| 6 | 52 | 7 | 405 | 0 | 11.21 | 1.51 | 87.28 | | | |
| 9 | 463 | 0 | 1 | 0 | 99.78 | 0.22 | | | | |
| 14 | 1 | 0 | 0 | 463 | 0.22 | | 99.78 | | | |
| 15 | 0 | 452 | 0 | 12 | | 97.41 | 2.59 | | | |
| 18 | 0 | 76 | 7 | 381 | | 16.38 | 1.51 | 82.11 | | |
| 21 | 0 | 295 | 0 | 169 | | 63.58 | 36.42 | | | |
| 22 | 8 | 0 | 456 | 0 | 1.72 | 98.28 | | | | |
| 27 | 3 | 451 | 0 | 10 | 0.65 | 97.20 | 2.16 | | | |
| 28 | 2 | 1 | 461 | 0 | 0.43 | 0.22 | 99.35 | | | |
| 29 | 1 | 0 | 463 | 0 | 0.22 | 99.78 | | | | |
| 30 | 3 | 341 | 1 | 119 | 0.65 | 73.49 | 0.22 | 25.65 | | |
| 31 | 0 | 10 | 0 | 454 | | 2.16 | 97.84 | | | |
| 33 | 2 | 5 | 455 | 2 | 0.43 | 1.08 | 98.06 | 0.43 | | |
| 36 | 51 | 0 | 410 | 3 | 10.99 | 88.36 | 0.65 | | | |
| 37 | 0 | 463 | 0 | 1 | | 99.78 | 0.22 | | | |

See also:

- *Investigating allele differences.*
- *Polymorphism analysis following isolate query.*

12.1.2 Codon usage

Select the alleles that you would like to include in the analysis. Again, variable length loci are limited to 200 sequences or fewer since these need to be aligned. Click ‘Codon’.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

Locus Explorer - Neisseria locus/sequence definitions

Please select locus for analysis:

Locus: adk Page will reload when changed

Select sequences Analysis functions

459 460 461 462 463 464

Polymorphic sites Display polymorphic site frequencies and sequence schematic

Codon Calculate G+C content and codon usage

Translate Translate DNA to peptide sequences

All None

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FetA | Options | Isolate Database

Codon Usage

adk

ORF used: 1

464 alleles included in analysis.

GC content

Coding: GC 52.47%
 1st letter: GC 65.31%
 2nd letter: GC 31.73%
 3rd letter: GC 60.37%

Codons

Fraction: Proportion of usage of a given codon among its redundant set (i.e. the set of codons which code for this codon's amino acid).
 Frequency: Usage of given codon per 1000 codons.

| Codon | Amino acid | Fraction | Frequency | Number |
|-------|------------|----------|-----------|--------|
| GCA | A | 0.262 | 17.353 | 1248 |
| GCC | A | 0.246 | 16.254 | 1169 |
| GCG | A | 0.389 | 25.751 | 1852 |
| GCT | A | 0.103 | 6.813 | 490 |
| TGC | C | 0.987 | 6.452 | 464 |
| TGT | C | 0.013 | 0.083 | 6 |
| GAC | D | 0.747 | 91.073 | 6550 |
| GAT | D | 0.253 | 30.812 | 2216 |
| GAA | E | 0.916 | 82.397 | 5926 |
| GAG | E | 0.084 | 7.564 | 544 |
| TTC | F | 0.594 | 15.295 | 1100 |
| TTT | F | 0.406 | 10.470 | 753 |
| GGA | G | 0.007 | 0.542 | 39 |
| GGC | G | 0.765 | 59.497 | 4279 |
| GGG | G | 0.001 | 0.042 | 3 |
| GGT | G | 0.227 | 17.659 | 1270 |
| CAC | H | 0.749 | 19.258 | 1385 |
| CAT | H | 0.251 | 6.438 | 463 |
| ATA | I | 0.001 | 0.083 | 6 |
| ATC | I | 0.698 | 40.788 | 2932 |

12.1.3 Aligned translations

If a DNA coding sequence locus is selected, an aligned translation can be produced.

Select the alleles that you would like to include in the analysis. Again, variable length loci are limited to 200 sequences or fewer since these need to be aligned. Click 'Translate'.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
 Download: Alleles | MLST profiles
 Links: Contents | Home | PorA | FetA | Options | Isolate Database

Locus Explorer - *Neisseria* locus/sequence definitions

Please select locus for analysis:

Locus: adk ▼ Page will reload when changed

Select sequences: 459 460 461 462 463 464 ▼

Analysis functions:

Polymorphic sites Display polymorphic site frequencies and sequence schematic

Codon Calculate G+C content and codon usage

Translate Translate DNA to peptide sequences

All None

An aligned amino acid sequence will be displayed.

```

433_1 .....S.....
434_1 .....L.....
435_1 .....T.....
436_1 .....
437_1 .....
438_1 .....
439_1 .....
440_1 .....K.....
441_1 .....
442_1 .....L.....
443_1 .....
444_1 .....
445_1 .....
446_1 .....
447_1 .....
448_1 .....
449_1 .....
450_1 .....
451_1 .....
452_1 .....K.....T.....
453_1 .....
454_1 .....IA.....T.....
455_1 .....
456_1 .....
457_1 .....G.....A...N.....P.....IAF.....T.....
458_1 .....
459_1 .....
460_1 .....S.....
461_1 .....
462_1 .....
463_1 .....
464_1 .....I.....
Consensus EAKKIIDEGLVDDIIIGMVKERIAQDDCKNGFLFDGFPRLAQAEAMVEAGVDLDVAVVEIDVPDSDVDRMSGRRVHLASGRTHVITYNPPKVEGKDD

```

| | 110 | 120 | 130 | 140 | 150 |
|-----|-------|-------|-------|-------|-------|
| 1_1 | | | | | |
| 2_1 | | | | | |
| 3_1 | | | | | |
| 4_1 | | | | | |
| 5_1 | | | | | |
| 6_1 | | | | | |
| 7_1 | | | | | |
| 8_1 | | | | | |

If there appear to be a lot of stop codons in the translation, it is possible that the `orf` value in the *locus definition* is not set correctly.

12.2 Field breakdown

The field breakdown plugin for isolate databases displays the frequency of each value for fields stored in the isolates table. *Allele and scheme field breakdowns* are handled by a different plugin.

The breakdown function can be selected for the whole database by clicking the ‘Single field’ link in the Breakdown section of the main contents page.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field**
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

Alternatively, a breakdown can be displayed of the dataset returned from a query by clicking the 'Fields' button in the Breakdown list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

| | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------|------------------------------|------------------------|----|------|----------------------------|------|------|------|
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 0021/84 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

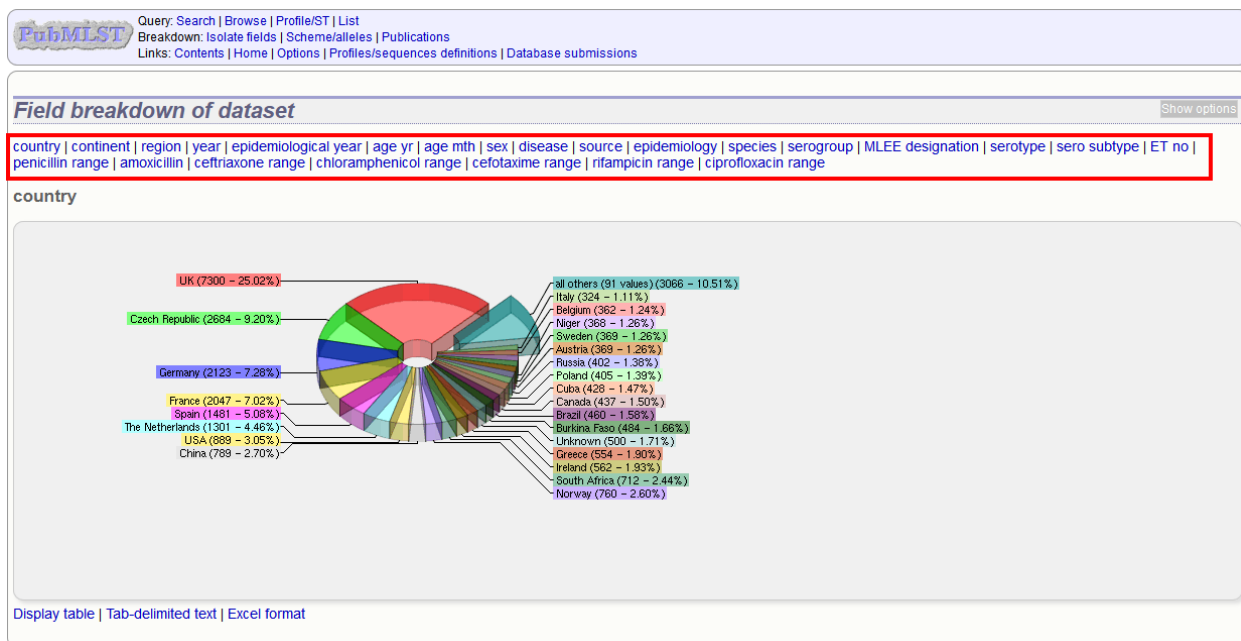
Breakdown: **Fields** | Two Field | Codons | Polymorphic sites | Combinations | Schemes/alleles | Publications | Sequence bin | Tag status

Analysis: BURST | Presence/Absence | Genome Comparator | BLAST

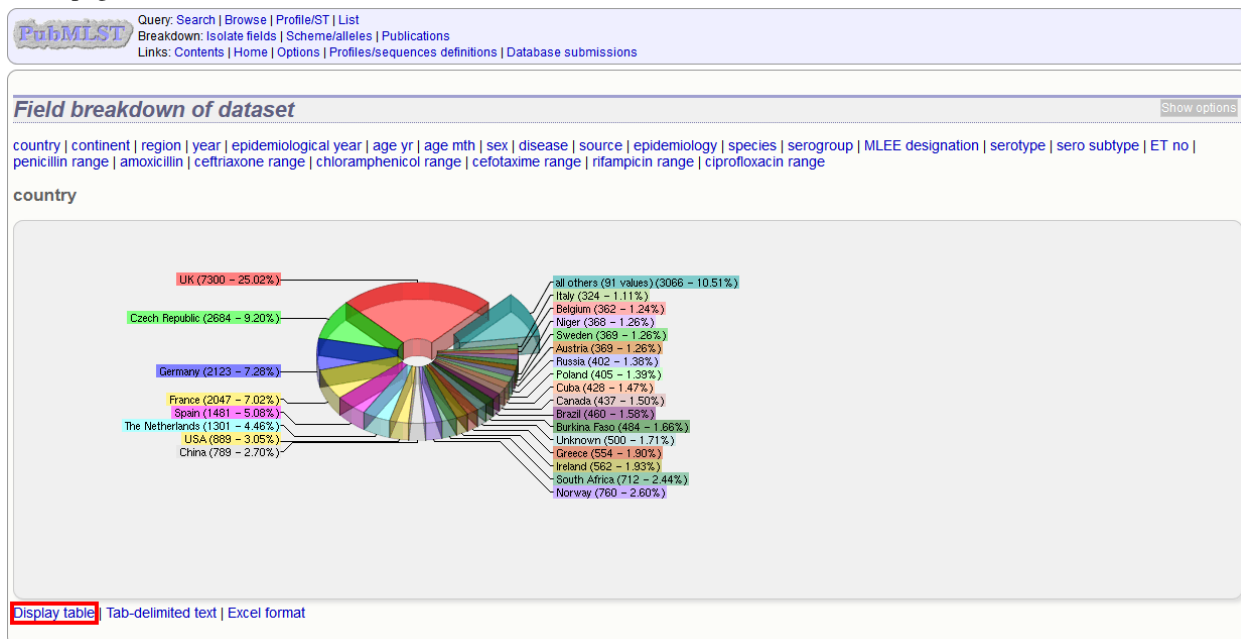
Export: Dataset | Contigs | Sequences

Page: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | > | Last

A series of charts will be displayed. Pick the field to display from the list at the top.



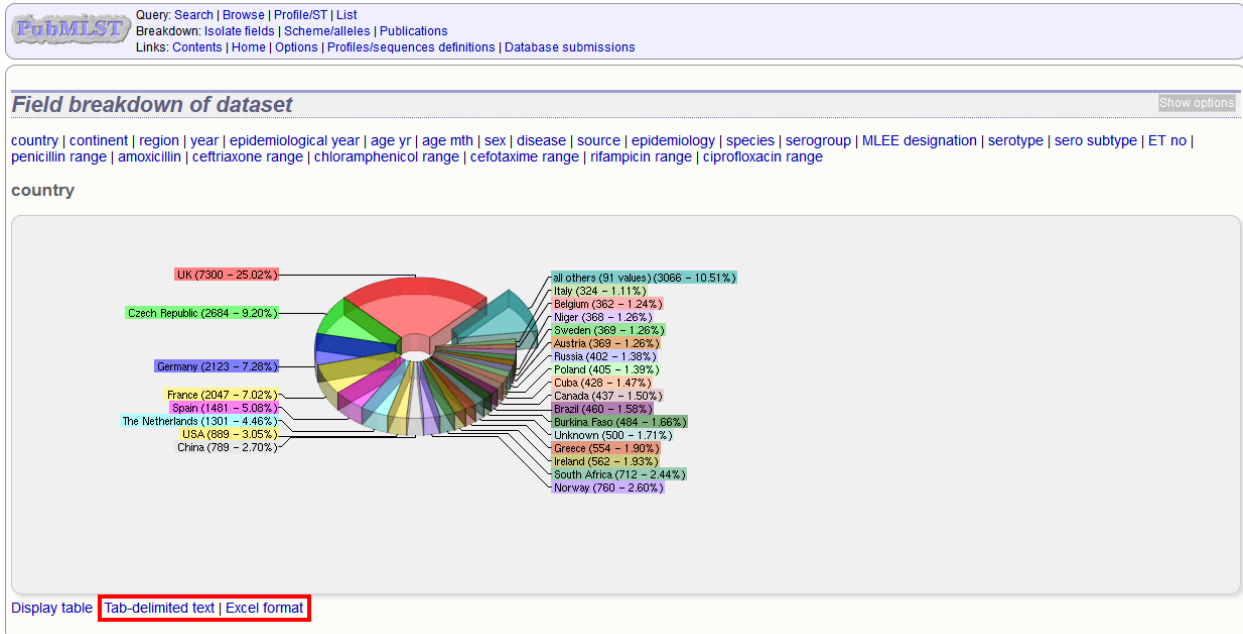
The values used to generate the chart can be displayed or extracted by clicking the ‘Display table’ link at the bottom of the page.



This displays a table that can be ordered by clicking the appropriate header.



The data can also be downloaded in tab-delimited text or Excel formats by clicking the appropriate links.



12.3 Two field breakdown

The two field breakdown plugin displays a table breaking down one field against another, e.g. breakdown of serogroup by year.

The analysis can be selected for the whole database by clicking the ‘Two field breakdown’ link on the main contents page.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- [Update history](#)
- [About BIGSdb](#)

Breakdown

- Single field
- Two field**
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

Alternatively, a two field breakdown can be displayed of the dataset returned from a query by clicking the ‘Two field’ button in the Breakdown list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

| | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------------------------------|------------------------------|------------------------|------|---------------|----------------------------|------|------|------|
| 5 | M0028220/ | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | SI-22 complex | | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 0021/84 | | | | | | | | | | | |
| 10 | 6748 | B73; Z1073 | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 11 | 129 | B92; Z1092 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 18-1 | 3 | F5-1 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 5-2 | 10 | F3-6 |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 7 | 16 | |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | 5-2 | 10 | F5-1 |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | 14 | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown:

Analysis:

Export:

Page:

Select the two fields you wish to breakdown and how you would like the values displayed (percentage/absolute values and totaling options).

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Two field breakdown of dataset Show options

Here you can create a table breaking down one field by another, e.g. breakdown of serogroup by year.

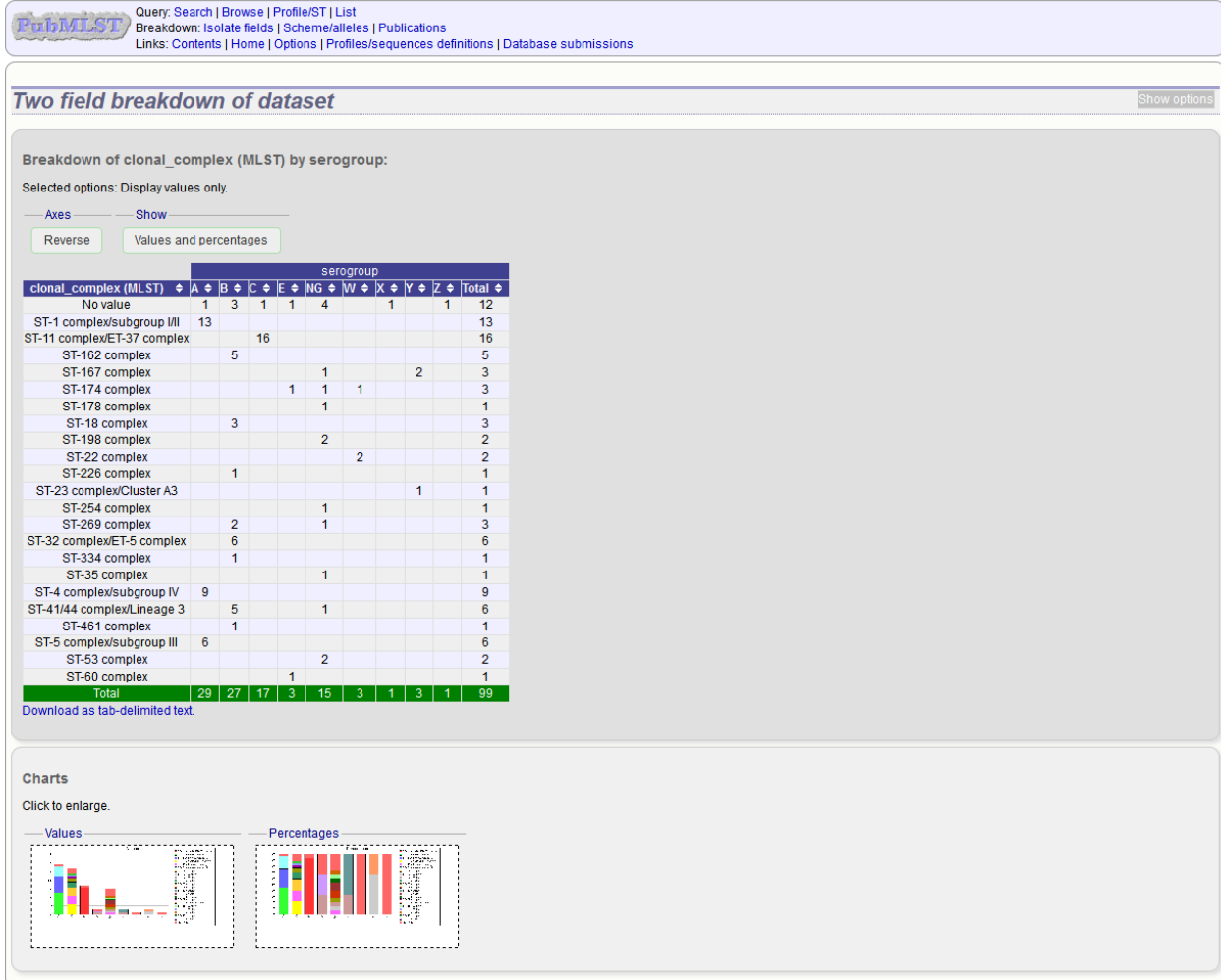
Select fields
Field 1: clonal complex (MLST)
Field 2: serogroup

Display
☒ values only
☐ values and percentages
☐ percentages only

Calculate percentages by
☒ dataset
☐ row
☐ column

Action
Reset Submit

Click submit. The breakdown will be displayed as a table. Bar charts will also be displayed provided the number of returned values for both fields are fewer than 30.



The table values can be exported in a format suitable for copying in to a spreadsheet by clicking 'Download as tab-delimited text' underneath the table.

12.4 Scheme and allele breakdown

The scheme and allele breakdown plugin displays the frequency of each allele and scheme field (e.g. ST or clonal complex).

The function can be selected for the whole database by clicking the ‘Scheme and allele breakdown’ link on the main contents page.

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles**
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

Alternatively, a breakdown can be displayed of the dataset returned from a query by clicking the ‘Schemes/alleles’ button in the Breakdown list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

| Isolate ID | Accession | Country | Year | Pathogen | Strain | ST | CC | PFGE | MLVA | Other | | |
|------------|-----------|-------------|----------------|----------|------------------------------|------------------------|----|------|----------------------------|-------|------|------|
| 1 | A4M1027 | B1; Z1001 | USA | 1937 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 5-2 | 10 | F1-5 |
| 2 | 120M | B35; Z1035 | Pakistan | 1967 | meningitis and septicaemia | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 3 | M00242905 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1099 | | 19 | 15 | |
| 4 | M1027 | B43; Z1043 | USA | 1937 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | | | |
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 002184 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown: Fields Two Field Codons Polymorphic sites Combinations **Schemes/alleles** Publications Sequence bin Tag status

Analysis: BURST Presence/Absence Genome Comparator BLAST

Export: Dataset Contigs Sequences

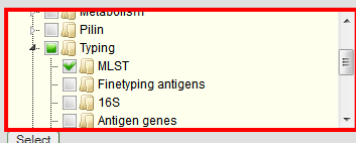
Page: 1 2 3 4 5 6 7 8 9 > Last

A scheme tree is shown. Select any combination of schemes to analyse.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Scheme field and allele breakdown of dataset Show options

Select schemes or groups of schemes within the tree. A breakdown of the individual fields and loci belonging to these schemes will then be performed.



Select

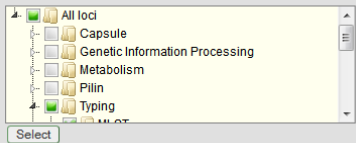
Click 'Select'.

A table showing the number of unique values for each locus and scheme field will be displayed.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Scheme field and allele breakdown of dataset Show options

Select schemes or groups of schemes within the tree. A breakdown of the individual fields and loci belonging to these schemes will then be performed.



Select

MLST

| Fields | | | Alleles | | | |
|----------------|---------------|---------------------------|---------|----------------|---------------------------|--------------------------|
| Field name | Unique values | Analyse | Locus | Unique alleles | Analyse | Download |
| ST | 554 | Breakdown | abcZ | 64 | Breakdown | Download |
| clonal complex | 43 | Breakdown | adk | 57 | Breakdown | Download |
| | | | aroE | 80 | Breakdown | Download |
| | | | fumC | 85 | Breakdown | Download |
| | | | gdh | 79 | Breakdown | Download |
| | | | pdhC | 83 | Breakdown | Download |
| | | | pgm | 68 | Breakdown | Download |

A detailed display of allele or field frequencies can be displayed by clicking the appropriate 'Breakdown' button.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Scheme field and allele breakdown of dataset

Show options

Select schemes or groups of schemes within the tree. A breakdown of the individual fields and loci belonging to these schemes will then be performed.

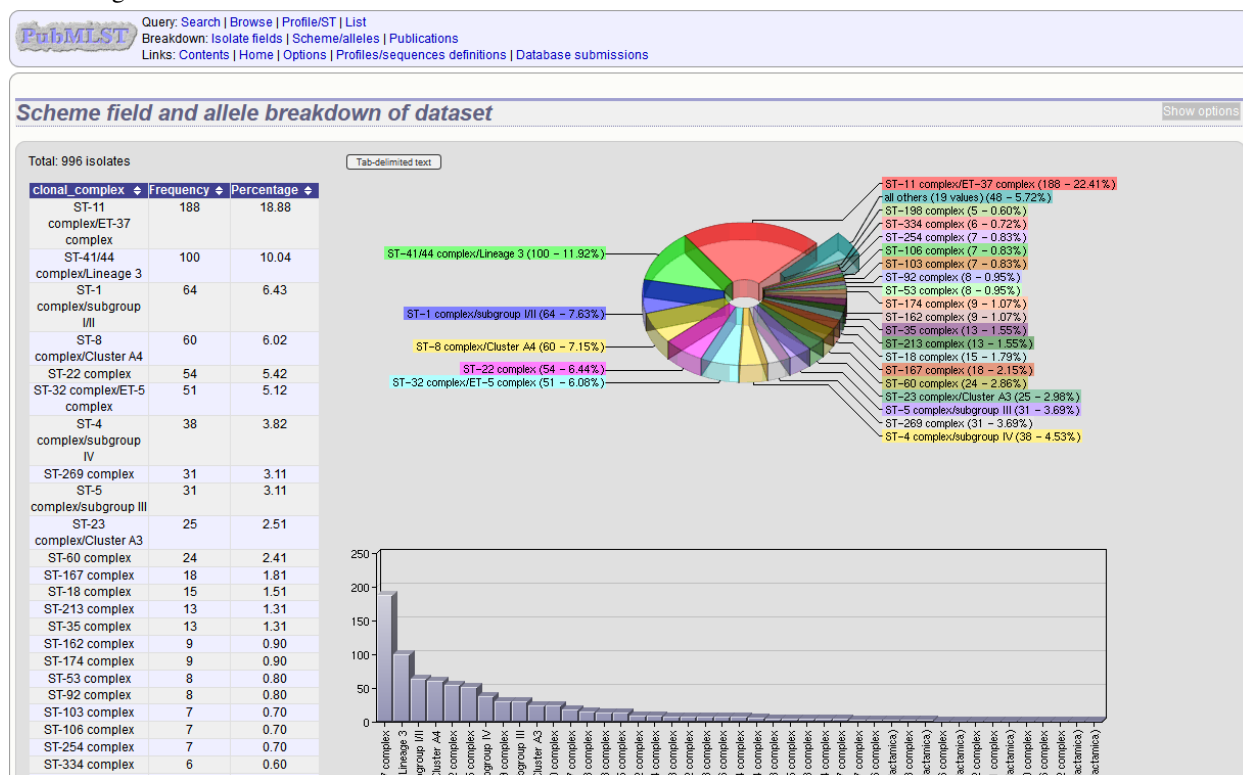
- All loci
- Capsule
- Genetic Information Processing
- Metabolism
- Pilin
- Typing

Select

MLST

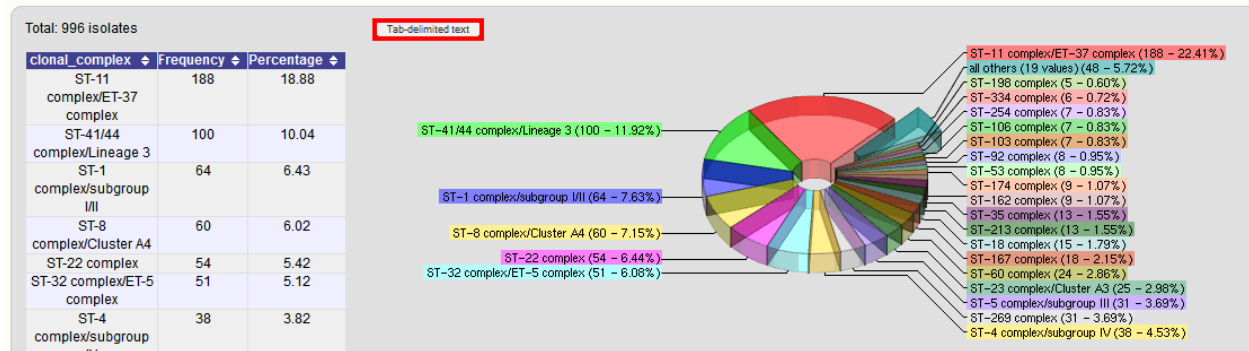
| Fields | | | Alleles | | | |
|----------------|---------------|-----------|---------|----------------|-----------|----------|
| Field name | Unique values | Analyse | Locus | Unique alleles | Analyse | Download |
| ST | 554 | Breakdown | abcZ | 64 | Breakdown | Download |
| clonal complex | 43 | Breakdown | adk | 57 | Breakdown | Download |
| | | | aroE | 80 | Breakdown | Download |
| | | | fumC | 85 | Breakdown | Download |
| | | | gdh | 79 | Breakdown | Download |
| | | | pdhC | 83 | Breakdown | Download |
| | | | pgm | 68 | Breakdown | Download |

The sorting of the table can be changed by clicking the appropriate header - this toggles between ascending and descending order.



The table values can be exported in a format suitable for copying in to a spreadsheet by clicking the 'Tab-delimited text' button.

Scheme field and allele breakdown of dataset



You can also download the sequences for alleles designated in the dataset for the loci belonging to the scheme by clicking the appropriate 'Download' button in the first results table.

PubMLST Query: Search | Browse | Profile/ST | List
 Breakdown: Isolate fields | Scheme/alleles | Publications
 Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Scheme field and allele breakdown of dataset

Select schemes or groups of schemes within the tree. A breakdown of the individual fields and loci belonging to these schemes will then be performed.

All loci
 Capsule
 Genetic Information Processing
 Metabolism
 Pilin
 Typing
 MLST

Select

| Fields | | | Alleles | | | |
|----------------|---------------|-----------|---------|----------------|-----------|----------|
| Field name | Unique values | Analyse | Locus | Unique alleles | Analyse | Download |
| ST | 554 | Breakdown | abcZ | 64 | Breakdown | Download |
| clonal complex | 43 | Breakdown | adk | 57 | Breakdown | Download |
| | | | aroE | 80 | Breakdown | Download |
| | | | fumC | 85 | Breakdown | Download |
| | | | gdh | 79 | Breakdown | Download |
| | | | pdhC | 83 | Breakdown | Download |
| | | | pgm | 68 | Breakdown | Download |

Sequences will be served in FASTA format in order of frequency.

```
>2
TTTGATACCGTTGCCGAAGGTTTGGGTGAAATTCGCGATTTATTGCGCCGTTACCAACGC
GTCGCCATGAGTTGGAACCGTTTCGGGTGAGGCTTTGTTGAAAGAACTCAACGAATTA
CAACTTGAAATCGAAGCGAAGGACGGCTGGAAGCTGGATGCGGCAGTCAAGCAGACTTTG
GGGGAACTCGGTTTCCCGGAAAACGAAAAATCGGCAACCTTTCCGGCGGTCAGAAAAAG
CGTGTCGCCTTGGCGCAGGCTTGGGTGCAGAAGCCCGACGTATTGCTGCTGGACGAACCG
ACCAACCATTGATATCGACGCGATTATTTGGCTGGAAAATCTGCTCAAAGCGTTTGAA
GGCAGCTTGGTTGTGATTACCCACGACCGCGCTTTTGGACAATATCGCCACGCGGATT
GTCGAACTCGATC
>1
TTTGATACTGTTGCCGAAGGTTTGGGCGAAATTCGCGATTTATTGCGCCGTTATCATCAT
GTCAGCCATGAGTTGGAATGTTTCGAGTGAGGCTTATTGAAAGAGCTCAACGAATTG
CAACTTGAGATCGAAGCGAAGGACGGCTGGAAGTTGGATGCGGCGGTGAAGCAGACTTTG
GGCGAACTCGGTTTCCCGGAAAACGAAAAATCGGCAACCTCTCCGGCGGTCAGAAAAAG
CGCGTCGCCTTGGCGCAGGCTTGGGTGCAGAAGCCCGACGTATTGCTGCTCGATGAACCG
ACCAACCATTGACATCGACGCGATTATTTGGTTGGAAAACCTGCTCAAAGCGTTTGAA
GGCAGCTTGGTTGTGATTACCCACGACCGCGCTTTTGGACAATATCGCCACGCGGATT
GTCGAACTCGATC
>4
```

```
TTTGATACCGTTGCCGAAGGTTTGGGCGAAATTCGTGATTATTGCGCCGTTATCATCAT
GTCAGCCATGAGTTGGAATGGTTCGAGTGAGGCTTTGTTGAAAGAACTCAACGAATTG
CAACTTGAAATCGAAGCGAAGGACGGCTGGAACCTGGATGCGGCAGTCAAGCAGACTTTG
GGGGAACTCGGTTTGCCGGAATGAAAAATCGGCAACCTTTCCGGCGGTGAGAAAAAG
CGCGTCGCCTTGGCTCAGGCTTGGGTGCAAAAGCCCGACGTATTGCTGCTGGACGAGCCG
ACCAACCATTGGATATCGACGCGATTATTTGGCTGGAAATCTGCTCAAAGCGTTTGAA
GGCAGCTTGGTTGTGATTACCCACGACCGCCGTTTTTTGGACAATATCGCCACGCGGATT
GTCGAACTCGATC
```

12.5 Sequence bin breakdown


The sequence bin breakdown plugin calculates statistics based on the number and length of contigs in the sequence bin as well as the number of loci tagged for an isolate record.

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


Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.


Query database


- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.


Option settings


- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.


Submissions

- Manage submissions


General information

- Isolates: 34221
- Last updated: 2015-07-02
- [Update history](#)
- [About BIGSdb](#)


Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- [Publications](#)
- [Sequence bin](#)**


Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats


Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST


Miscellaneous

- Description of database fields

Alternatively, it can be accessed following a query by clicking the ‘Sequence bin’ button in the Breakdown list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

| | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------|------------------------------|------------------------|----|------|----------------------------|------|----|------|
| 2 | 72UM | B35; Z1035 | Pakistan | 1997 | meningitis and septicemia | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F5-1 |
| 3 | M00242905 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1099 | | | 19 | 15 |
| 4 | M1027 | B43; Z1043 | USA | 1937 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | | | |
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | | 7 | 16 |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | | 20 | 9 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | | 14 |
| 9 | 0021/84 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | | 7 | 16 |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | | 14 |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | | 7 | 13-1 |
| 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 | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown: Fields Two Field Codons Polymorphic sites Combinations Schemes/alleles Publications **Sequence bin** Tag status

Analysis: BURST Presence/Absence Genome Comparator BLAST

Export: Dataset Contigs Sequences

Page: 1 2 3 4 > Last

Select the isolate records to analyse - these will be pre-selected if you accessed the plugin following a query. You can also select loci and/or schemes which will be used to calculate the totals and percentages of loci designated and tagged. This may be useful as a guide to assembly quality if you use a scheme of core loci where a good assembly would be expected to include all member loci. To determine the total of all loci designated or tagged, click 'All loci' in the scheme tree.

There is also an option to determine the mean G+C content of the sequence bin of each isolate.

Click submit.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Breakdown of sequence bin contig properties

Please select the required isolate ids for comparison - use Ctrl or Shift to make multiple selections. Select loci/schemes to use for calculating percentage of alleles designated or tagged.

Isolates

- 64) 254
- 67) S5611
- 82) 11-004
- 84) IAL2229
- 90) CN100
- 120) F4698
- 128) F6124
- 160) 1014

Loci

- *16S_rDNA
- 16S_rRNA (SSU_rRNA)
- *23S_rRNA
- abcZ (NEIS1015)
- aceF (NEIS1279)
- ackA2 (NEIS1727)
- acnA (NEIS1729)

Schemes

- All loci**
- Capsule
- Genetic Information Proc...
- Genomic Islands
- Lineage Schemes
- Metabolism
- Pilin

Options

☐ Calculate %GC

Filter by

Sequence method:

Project:

Experiment:

Action

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.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions

Help

Breakdown of sequence bin contig properties

Please select the required isolate ids for comparison - use Ctrl or Shift to make multiple selections. Select loci/schemes to use for calculating percentage of alleles designated or tagged.

Isolates: 64) 254, 67) S5611, 82) 11-004, 84) IAL2229, 90) CN100, 120) F4698, 128) F6124, 160) 1014
Loci: 16S_rDNA, 16S_rRNA (SSU_rRNA), 23S_rRNA, abcZ, abcZ (NEIS1015), aceF (NEIS1279), ackA2 (NEIS1727), acnA (NEIS1729)
Schemes: All loci, Capsule, Capsule Region A, Capsule Region A, Capsule Region A, Capsule Region A
Options: ☐ Calculate %GC

Filter by: Sequence method: Project: Experiment: Action: Reset Submit

Loci selected: 2397

| Isolate id | Isolate | Contigs | Total length | Min | Max | Mean | σ | N50 contig number | N50 contig length (L50) | N90 contig number | N90 contig length (L90) | N95 contig number | N95 contig length (L95) | Alleles designated | % Alleles designated | Loci tagged | % Loci tagged | Sequence bin |
|------------|---------|---------|--------------|-----|--------|-------|----------|-------------------|-------------------------|-------------------|-------------------------|-------------------|-------------------------|--------------------|----------------------|-------------|---------------|--------------|
| 64 | 254 | 251 | 2054183 | 257 | 69510 | 8184 | 9379 | 42 | 15763 | 131 | 4488 | 162 | 2543 | 1817 | 75.8 | 1892 | 78.9 | Display |
| 67 | S5611 | 194 | 2084213 | 255 | 103457 | 10744 | 14516 | 28 | 22332 | 94 | 6156 | 115 | 3531 | 1852 | 77.3 | 1898 | 79.2 | Display |
| 82 | 11-004 | 258 | 2055227 | 251 | 47478 | 7966 | 8518 | 45 | 13854 | 139 | 4751 | 167 | 2921 | 1830 | 76.3 | 1886 | 78.7 | Display |
| 84 | IAL2229 | 188 | 2051679 | 256 | 96122 | 10914 | 13975 | 28 | 21515 | 93 | 6282 | 115 | 3629 | 1848 | 77.1 | 1896 | 79.1 | Display |
| 90 | CN100 | 236 | 2118544 | 261 | 73069 | 8977 | 11101 | 36 | 19065 | 119 | 4570 | 149 | 2839 | 1848 | 77.1 | 1898 | 79.2 | Display |

[Download in tab-delimited text format](#)
[Download in Excel format](#)

Click on the following charts to enlarge

Number of contigs: Overall mean: 225.4; σ : 32.5
Total length: Overall mean: 2072769.2; σ : 28828.2
Mean contig length: Overall mean: 9357.0; σ : 1396.7
Contig lengths: Overall mean: 9196.0; σ : 11481.9

Download lengths

Sequence bin records can also be accessed by clicking the 'Display' button for each row of the table.

Filter by: Sequence method: Project: Experiment: Action: Reset Submit

Loci selected: 2397

| Isolate id | Isolate | Contigs | Total length | Min | Max | Mean | σ | N50 contig number | N50 contig length (L50) | N90 contig number | N90 contig length (L90) | N95 contig number | N95 contig length (L95) | Alleles designated | % Alleles designated | Loci tagged | % Loci tagged | Sequence bin |
|------------|---------|---------|--------------|-----|--------|-------|----------|-------------------|-------------------------|-------------------|-------------------------|-------------------|-------------------------|--------------------|----------------------|-------------|---------------|--------------|
| 64 | 254 | 251 | 2054183 | 257 | 69510 | 8184 | 9379 | 42 | 15763 | 131 | 4488 | 162 | 2543 | 1817 | 75.8 | 1892 | 78.9 | Display |
| 67 | S5611 | 194 | 2084213 | 255 | 103457 | 10744 | 14516 | 28 | 22332 | 94 | 6156 | 115 | 3531 | 1852 | 77.3 | 1898 | 79.2 | Display |
| 82 | 11-004 | 258 | 2055227 | 251 | 47478 | 7966 | 8518 | 45 | 13854 | 139 | 4751 | 167 | 2921 | 1830 | 76.3 | 1886 | 78.7 | Display |
| 84 | IAL2229 | 188 | 2051679 | 256 | 96122 | 10914 | 13975 | 28 | 21515 | 93 | 6282 | 115 | 3629 | 1848 | 77.1 | 1896 | 79.1 | Display |
| 90 | CN100 | 236 | 2118544 | 261 | 73069 | 8977 | 11101 | 36 | 19065 | 119 | 4570 | 149 | 2839 | 1848 | 77.1 | 1898 | 79.2 | Display |

[Download in tab-delimited text format](#)
[Download in Excel format](#)

Click on the following charts to enlarge

12.6.1 Analysis using defined loci

Select the isolate genomes that you wish to analyse and then either the loci from the list or a set of schemes. Press submit.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Genome Comparator - *Neisseria* PubMLST

Please select the required isolate ids and loci for comparison - use ctrl or shift to make multiple selections. 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

- 658) 890326
- 659) A22
- 660) 71194
- 661) 860800
- 662) 2837
- 663) 2839
- 664) 2838
- 665) 2845

Loci

- 16S_rDNA
- 16S_rRNA (SSU_rRNA)
- 23S_rRNA
- abcZ
- abcZ (NEIS1015)
- aceF (NEIS1279)
- ackA2 (NEIS1727)
- ackA (NEIS1729)

Include in identifiers

- isolate
- country
- region
- year
- epidemiological year
- age yr
- age mth
- sex
- disease
- source

Schemes

- Genomic islands
- Lineage Schemes
- Metabolism
- Pilin
- Typing
- Finotyping antigens

Reference genome

Enter accession number:

or choose annotated genome:

or upload Genbank/EMBL/FASTA file: No file selected.

Parameters / options

Min % identity:

Min % alignment:

BLASTN word size:

☒ Use TBLASTX

☒ Use tagged designations if available

☐ Disable HTML output

Distance matrix calculation

With incomplete loci:

- ☒ Completely exclude from analysis
- ☐ Treat as distinct allele
- ☐ Ignore in pairwise comparison

☒ Exclude paralogous loci

- ☐ paralogous in all isolates
- ☐ paralogous in any isolate

Alignments

☐ Produce alignments

☒ Include ref sequences in alignment

☐ Align all loci (not only variable)

Aligner:

Core genome analysis

Core threshold (%):

☐ Calculate mean distances

Filter by

Sequence method:

Project:

Experiment:

Action

The job will be submitted to the job queue and will start running shortly. Click the link to follow the job progress and view the output.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Genome Comparator - *Neisseria* PubMLST

This analysis has been submitted to the job queue.

Please be aware that this job may take a long time depending on the number of comparisons and how busy the server is.

[Follow the progress of this job and view the output.](#)

There will be a series of tables displaying variable loci, colour-coded to indicate allelic differences. Finally, there will be links to a distance matrix which can be loaded in to SplitsTree for further analysis and to a NeighborNet chart showing relatedness of isolates. Due to processing constraints on the web server, this NeighborNet is only calculated if 200 or fewer genomes are selected for analysis, but this can be generated in the stand-alone version of SplitsTree using the distance matrix if required.



Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Job status viewer

Status

| | |
|--------------|-------------------------------|
| Job id: | BIGSdb_27748_1405410063_94241 |
| Submit time: | 2014-07-15 08:41:03 |
| Status: | finished |
| Start time: | 2014-07-15 08:41:05 |
| Progress: | 100% |
| Stop time: | 2014-07-15 08:41:22 |
| Total time: | 17 seconds |

Output

Analysis against defined loci

Allele numbers are used where these have been defined, otherwise sequences will be marked as 'New#1', 'New#2' etc. Missing alleles are marked as 'X'. Truncated alleles (located at end of contig) are marked as 'T'.

| Locus | 644 (L93/4286) | 662 (2837) | 663 (2839) | 664 (2838) | 665 (2845) | 666 (2843) | 667 (2842) | 669 (2846) | 670 (2840) | 671 (2844) | 672 (2847) | 698 (FAM18) |
|-------|----------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| abcZ | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 7 | 2 | 2 | 2 | 2 |
| adk | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| aroE | 4 | 19 | 4 | 4 | 4 | 4 | 4 | 4 | 19 | 19 | 19 | 4 |
| fumC | 3 | 3 | 3 | 3 | 24 | 23 | 23 | 3 | 3 | 3 | 3 | 3 |
| gdh | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| pdhC | 4 | 4 | 4 | 4 | 4 | 6 | 6 | 4 | 4 | 4 | 4 | 4 |
| pgm | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |

Loci with sequence differences among isolates:

Variable loci: 4

| Locus | 644 (L93/4286) | 662 (2837) | 663 (2839) | 664 (2838) | 665 (2845) | 666 (2843) | 667 (2842) | 669 (2846) | 670 (2840) | 671 (2844) | 672 (2847) | 698 (FAM18) |
|-------|----------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| abcZ | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 7 | 2 | 2 | 2 | 2 |
| aroE | 4 | 19 | 4 | 4 | 4 | 4 | 4 | 4 | 19 | 19 | 19 | 4 |
| fumC | 3 | 3 | 3 | 3 | 24 | 23 | 23 | 3 | 3 | 3 | 3 | 3 |
| pdhC | 4 | 4 | 4 | 4 | 4 | 6 | 6 | 4 | 4 | 4 | 4 | 4 |

Exactly matching loci

These loci are identical in all isolates.

Matches: 3

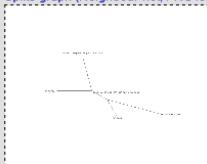
| Locus | 644 (L93/4286) | 662 (2837) | 663 (2839) | 664 (2838) | 665 (2845) | 666 (2843) | 667 (2842) | 669 (2846) | 670 (2840) | 671 (2844) | 672 (2847) | 698 (FAM18) |
|-------|----------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| adk | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| gdh | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| pgm | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |

Unique strains

Unique strains: 5

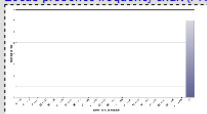
| Strain 1 | Strain 2 | Strain 3 | Strain 4 | Strain 5 |
|----------------|------------|------------|------------|------------|
| 644 (L93/4286) | 662 (2837) | 666 (2843) | 665 (2845) | 669 (2846) |
| 663 (2839) | 670 (2840) | 667 (2842) | | |
| 664 (2838) | 671 (2844) | | | |
| 698 (FAM18) | 672 (2847) | | | |

- [Text output file](#)
- [Excel format](#)
- [Distance matrix \(Nexus format\)](#) - Suitable for loading in to [SplitsTree](#). Distances between taxa are calculated as the number of loci with different allele sequences
- [Splits graph \(Neighbour-net, PNG format\)](#)



(click to enlarge)

- [Splits graph \(Neighbour-net, SVG format\)](#) - This can be edited in [Inkscape](#) or other vector graphics editors
- [Locus presence frequency](#)
- [Locus presence frequency chart \(PNG format\)](#)



(click to enlarge)

- [Tar file containing output files](#)

Please note that job results will remain on the server for 7 days.

12.6.2 Analysis using annotated reference genome

Select the isolate genomes that you wish to analyse and then either enter a Genbank accession number for the reference genome, or select from the list of reference genomes (this list will only be present if the administrator has *set it up*). Selecting reference genomes will hide the locus and scheme selection forms.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Genome Comparator - Neisseria PubMLST

Please select the required isolate ids and loci for comparison - use ctrl or shift to make multiple selections. 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

658) 890326
659) A22
660) 71/94
661) 860800
662) 2837
663) 2839
664) 2838
665) 2845

Include in identifiers

isolate
country
region
year
epidemiological year
age yr
age mth
sex
disease
source

Reference genome

Enter accession number:
or choose annotated genome:
FAM18 (Nm)
or upload Genbank/EMBL/FASTA file:
Browse... No file selected.

Parameters / options

Min % identity: 70
Min % alignment: 50
BLASTN word size: 15
☒ Use TBLASTX
☒ Use tagged designations if available
☐ Disable HTML output

Distance matrix calculation

With incomplete loci:
☒ Completely exclude from analysis
☐ Treat as distinct allele
☐ Ignore in pairwise comparison
☒ Exclude paralogous loci
☐ paralogous in all isolates
☐ paralogous in any isolate

Alignments

☐ Produce alignments
☒ Include ref sequences in alignment
☐ Align all loci (not only variable)
Aligner: MAFFT

Core genome analysis

Core threshold (%): 90
☐ Calculate mean distances

Filter by

Sequence method:
Project:
Experiment:

Action

Reset Submit

Output is similar to when comparing against defined loci, but this time every coding sequence in the annotated reference will be BLASTed against the selected genomes. Because allele designations are not defined, the allele found in the reference genome is designated allele 1, the next different sequence is allele 2 etc.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Job status viewer

Status

Job id: BIGSdb_21911_1405410488_8452
Submit time: 2014-07-15 08:48:08
Status: finished
Start time: 2014-07-15 08:48:34
Progress: 100%
Stop time: 2014-07-15 09:12:58
Total time: 24 minutes and 24 seconds

Output

Analysis by reference genome

| | |
|----------------|---|
| accession | AM421808 |
| version | 1 |
| type | dna |
| length | 2194961 |
| description | Neisseria meningitidis serogroup C FAM18 complete genome. |
| coding regions | 1975 |

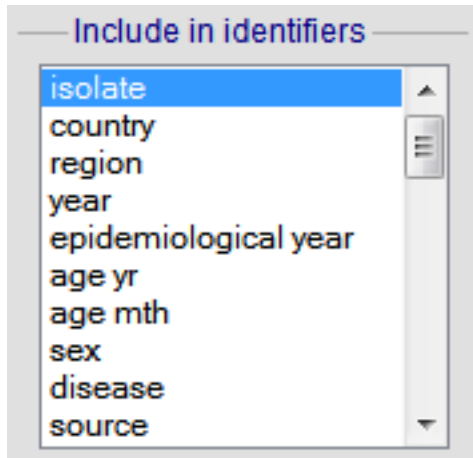
All loci

Each unique allele is defined a number starting at 1. Missing alleles are marked as 'X'. Truncated alleles (located at end of contig) are marked as 'T'.

| Locus | Product | Sequence length | Genome position | Reference genome | 644 (L93/4286) | 662 (2837) | 663 (2839) | 664 (2838) | 665 (2845) | 666 (2843) | 667 (2842) | 669 (2846) | 670 (2840) | 671 (2844) | 672 (2847) | 698 (FAM18) |
|-----------------------|---|-----------------|-----------------|------------------|----------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| lpxC envA NMC0001 | UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase | 924 | 1261 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| pilS1 NMC0002 | pilin (fragment) | 291 | 3341 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| pilS2 NMC0003 | truncated pilin | 366 | 3675 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |
| tpb NMC0004 | peptidyl-prolyl cis-trans isomerase | 330 | 4069 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | T | 2 | 2 | 2 | 1 |
| NMC0005 | putative membrane protein | 219 | 4476 | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 5 | 3 | 1 |
| NMC0006 | putative glycerate dehydrogenase | 954 | 4816 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |
| metG NMC0007 | methionyl-tRNA synthetase | 2058 | 5843 | 1 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |
| glmS NMC0008 | glucosamine-fructose-6-phosphate aminotransferase [isomerizing] | 1839 | 8016 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |

12.6.3 Include in identifiers fieldset

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



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

12.6.4 Reference genome fieldset

This section allows you to choose a reference genome to use as the source of comparator sequences.

There are three possibilities here:

1. Enter accession number - Enter a Genbank accession number of an annotated reference and Genome Comparator will automatically retrieve this from Genbank.
2. Select from list - The administrator may have selected some genomes to offer for comparison. If these are present, simply select from the list.
3. Upload genome - Click 'Browse' and upload your own reference. This can either be in Genbank, EMBL or FASTA format. Ensure that the filename ends in the appropriate file extension (.gb, .embl, .fas) so that it is recognized.

12.6.5 Parameters/options fieldset

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

Parameters / options

Min % identity: 70

Min % alignment: 50

BLASTN word size: 15

☐ Use TBLASTX

☒ Use tagged designations if available

☐ Disable HTML output

- **Min % identity** - This sets the threshold identity that a matching sequence has to be in order to be considered (default: 70%). Only the best match is used.
- **Min % alignment** - This sets the percentage of the length of reference allele sequence that the alignment has to cover in order to be considered (default: 50%).
- **BLASTN word size** - This is the length of the initial identical match that BLAST requires before extending a match (default: 20). Increasing this value improves speed at the expense of sensitivity. The default value gives good results in most cases. The default setting used to be 15 but the new default of 20 is almost as good (there was 1 difference among 2000 loci in a test run) but the analysis runs twice as fast.
- **Use TBLASTX** - This compares the six-frame translation of your nucleotide query sequence against the six-frame translation of the contig sequences. Sequences will be classed as identical if they result in the same translated sequence even if the nucleotide sequence is different. This is significantly slower than using BLASTN.

Additionally, two other options are available in this fieldset:

- **Use tagged designations** - When analysing using defined loci, Genome Comparator can use the designations stored within the database (this is the default). This is much quicker since it doesn't need to run BLAST against these sequences. If a designation is missing, BLAST will be run for that locus anyway.
- **Disable HTML output** - If running Genome Comparator against a large number of genomes, the resulting table may get so large that your web browser struggles to render it properly and may use up too much memory on your computer. Clicking this button prevents this output - this output is not required for further analysis since everything present in it is also generated in Excel format at the end. HTML output is automatically disabled when more than 150 genomes are analysed.

12.6.6 Distance matrix calculation fieldset

This section provides options for the treatment of incomplete and paralogous loci when generating the distance matrix.

Distance matrix calculation

With incomplete loci:

☒ Completely exclude from analysis

☐ Treat as distinct allele

☐ Ignore in pairwise comparison

☒ Exclude paralogous loci

☒ paralogous in all isolates

☐ paralogous in any isolate

For incomplete loci, i.e. those that continue beyond the end of a contig so are incomplete you can:

- Completely exclude from analysis - Any locus that is incomplete in at least one isolate will be removed from the analysis completely. Using this option means that if there is one bad genome with a lot of incomplete sequences in your analysis, a large proportion of the loci may not be used to calculate distances.
- Treat as a distinct allele - This treats all incomplete sequences as a specific allele 'I'. This varies from any other allele, but all incomplete sequences will be treated as though they were identical.
- Ignore in pairwise comparison (default) - This is probably the best option. In this case, incomplete alleles are only excluded from the analysis when comparing the particular isolate that has it. Other isolates with different alleles will be properly included. The effect of this option will be to shorten the distances of isolates with poorly sequenced genomes with the others.

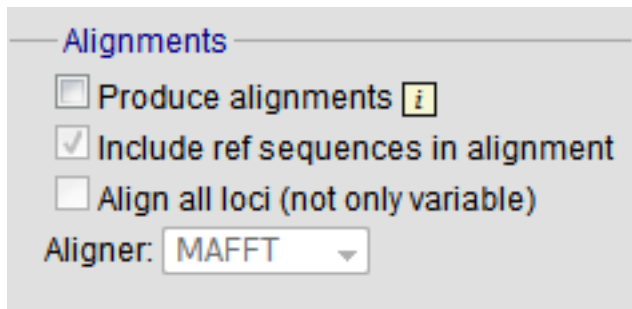
Paralogous loci, i.e. those with multiple good matches, can be excluded from the analysis (default). This is the safest option since there is no guarantee that differences seen between isolates at paralogous loci are real if the alternative matches are equally good. NB: Loci are also only classed as paralogous when the alternative matches identify different sequences, otherwise multiple contigs of the same sequence region would result in false positives.

When paralogous loci are excluded, there are two further options:


- Exclude when paralogous in all isolates (default). Loci are only classed as paralogous when there are multiple hits in every genome (except if a genome is missing the locus entirely, in which case that genome is ignored in the calculation). This is generally the option that you will want to use with the default BLAST parameters since you can often expect multiple hits even when loci are not paralogous if you have used relaxed thresholds.
- Exclude when paralogous in any isolate. Unless you use stringent BLAST thresholds, this is likely to overestimate the number of paralogous loci, but may be useful if you are specifically looking for them.

12.6.7 Alignments fieldset

This section enables you to choose to produce alignments of the sequences identified.



Alignments

☐ Produce alignments 

☒ Include ref sequences in alignment

☐ Align all loci (not only variable)

Aligner: MAFFT

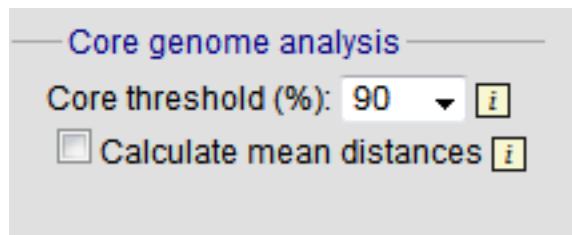
Available options are:

- 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 approximately twice as long to run.
- Include ref sequences in alignment - When doing analysis using an annotated reference, selecting this will include the reference sequence in the alignment files.
- Align all loci - By default, only loci that vary among the isolates are aligned. You may however wish to align all if you would like the resultant XMFA and FASTA files to include all coding sequences.
- Aligner - There are currently two choices of alignment algorithm (provided they have both been installed)
 - MAFFT (default) - This is the preferred option as it is significantly quicker than MUSCLE, uses less memory, and produces comparable results.

- MUSCLE - This was originally the only choice. It is still included to enable previous analyses to be re-run and compared but it is recommended that MAFFT is used otherwise.

12.6.8 Core genome analysis fieldset

This section enables you to modify the inclusion threshold used to calculate whether or not a locus is part of the core genome (of the dataset).

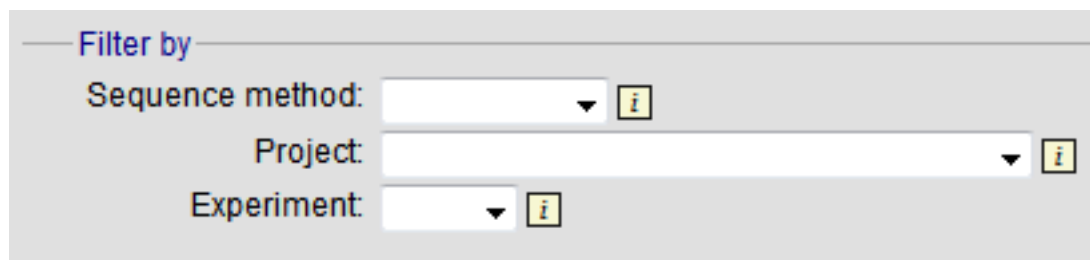


The default setting of 90% means that a locus is counted as core if it appears within 90% or more of the genomes in the dataset.

There is also an option to calculate the mean distance among sequences of the loci. Selecting this will also select the option to produce alignments.

12.6.9 Filter fieldset

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



Available options are:

- Sequence method - Choose to only analyse contigs that have been generated using a particular method. This depends on the method being set when the contigs were uploaded.
- Project - Only include isolates belonging to the chosen project. This enables you to select all isolates and filter to a project.
- Experiment - Contig files can belong to an experiment. How this is used can vary between databases, but this enables you to only include contigs from a particular experiment.

12.6.10 Understanding the output

Distance matrix

The distance matrix is simply a count of the number of loci that differ between each pair of isolates. It is generated in NEXUS format which can be used as the input file for [SplitsTree](#). This can be used to generate NeighborNet, Split decomposition graphs and trees offline. If 200 isolates or fewer are included in the analysis, a Neighbor network is automatically generated from this distance matrix.

Unique strains

The table of unique strains is a list of isolates that are identical at every locus. Every isolate is likely to be classed as unique if a whole genome analysis is performed, but with a constrained set of loci, such as those for MLST, this will group isolates that are indistinguishable at that level of resolution.

12.7 BLAST

The BLAST plugin enables you to BLAST a sequence against any of the genomes in the database, displaying a table of matches and extracting matching sequences.

The function can be accessed by selecting the 'BLAST' link on the Analysis section of the main contents page.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST**

Miscellaneous

- Description of database fields

Alternatively, it can be accessed following a query by clicking the 'BLAST' button in the Analysis list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

| | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------|------------------------------|------------------------|----|------|----------------------------|------|------|------|
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 0021/84 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown: Fields Two Field Codons Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status

Analysis: BURST Presence/Absence Genome Comparator **BLAST**

Export: Dataset Contigs Sequences

Page: 1 2 3 4 > Last

Select the isolate records to analyse - these will be pre-selected if you accessed the plugin following a query. Paste in a sequence to query - this be either a DNA or peptide sequence.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Toggle: [i]

BLAST

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 sequences can be queried.

Isolates
 64) 254
 67) S5611
 82) 11-004
 84) IAL2229
 90) CN100
 120) F4698
 128) F6124
 160) 1014

Paste sequence
 GAAGCCTTGGGCGACGCGATGCCGCCGTTGCCGCTTCGGGCGATTGAAAACGCTGGCG
 GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCGCGCTGCGGT
 TTGGGCGAAATCAAATCCCCGAAAACGAGCGGGTTCGTCCATCATGCCGGGCAAGTC
 AACCGGACCCAAATCGGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGAGCTT
 ACCATCGGCATGGCGGGCGGTGGGCAATTCGAGCTGAACGCTATATGCCGTTATC
 GCTTACAACCTCTTGCAATCCATCCGCTGTTGGGCGACGCGTGAACAGCTTCAACGAA
 CACTGCGCGCTGGGATCGAACCGTACCGGAAAAAATCGACTATTCTCGACCATTC
 CTGATGCTGGTACTCGGTTAAACGTAATAATCGGCTACGAAAC

Include in results table
 country
 region
 year
 epidemiological year
 age yr
 age mth
 sex
 disease
 source
 epidemiology

Parameters
 BLASTN word size: 11
 BLASTN scoring: reward:2, penalty:-3, gap open:5, gap extend:2
 Hits per isolate: 1
 Flanking length (bp): 100
☐ Use TBLASTX

Options
☐ Show 0% matches in table

Restrict included sequences by
 Sequence method:
 Project:
 Experiment:

Action

Click submit.

A table of BLAST results will be displayed.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Toggle: [i]

BLAST

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 sequences can be queried.

Isolates
 1) A4/M1027
 2) 120M
 7) 7891
 10) 6748
 11) 129
 13) 139M
 19) S3131
 24) S4355

Paste sequence
 GAAGCCTTGGGCGACGCGATGCCGCCGTTGCCGCTTCGGGCGATTGAAAACGCTGGCG
 GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCGCGCTGCGGT
 TTGGGCGAAATCAAATCCCCGAAAACGAGCGGGTTCGTCCATCATGCCGGGCAAGTC
 AACCGGACCCAAATCGGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGAGCTT
 ACCATCGGCATGGCGGGCGGTGGGCAATTCGAGCTGAACGCTATATGCCGTTATC
 GCTTACAACCTCTTGCAATCCATCCGCTGTTGGGCGACGCGTGAACAGCTTCAACGAA
 CACTGCGCGCTGGGATCGAACCGTACCGGAAAAAATCGACTATTCTCGACCATTC
 CTGATGCTGGTACTCGGTTAAACGTAATAATCGGCTACGAAAC

Include in results table
 country
 region
 year
 epidemiological year
 age yr
 age mth
 sex
 disease
 source
 epidemiology

Parameters
 BLASTN word size: 11
 BLASTN scoring: reward:2, penalty:-3, gap open:5, gap extend:2
 Hits per isolate: 1
 Flanking length (bp): 100
☐ Use TBLASTX

Options
☐ Show 0% matches in table

Restrict included sequences by
 Sequence method:
 Project:
 Experiment:

Action

| Isolate id | Isolate | % identity | Alignment length | Mismatches | Gaps | Seqbin id | Start | End | Orientation | E-value | Bit score |
|------------|----------|------------|------------------|------------|------|-----------|---------|---------|-------------|---------|-----------|
| 1 | A4/M1027 | 98.49 | 465 | 7 | 0 | 180177 | 19444 | 19908 | extract | 0.0 | 807 |
| 2 | 120M | 98.49 | 465 | 7 | 0 | 180583 | 4782 | 5246 | extract | 0.0 | 807 |
| 7 | 7891 | 98.49 | 465 | 7 | 0 | 180965 | 19869 | 20333 | extract | 0.0 | 807 |
| 10 | 6748 | 98.49 | 465 | 7 | 0 | 181186 | 19181 | 19645 | extract | 0.0 | 807 |
| 11 | 129 | 98.49 | 465 | 7 | 0 | 181867 | 35889 | 36353 | extract | 0.0 | 807 |
| 13 | 139M | 98.49 | 465 | 7 | 0 | 182004 | 36775 | 37239 | extract | 0.0 | 807 |
| 19 | S3131 | 98.49 | 465 | 7 | 0 | 182318 | 19090 | 19554 | extract | 0.0 | 807 |
| 24 | S4355 | 98.49 | 465 | 7 | 0 | 215673 | 4534 | 4998 | extract | 0.0 | 807 |
| 30 | 14 | 97.42 | 465 | 12 | 0 | 8 | 1363524 | 1363988 | extract | 0.0 | 785 |
| 31 | 10 | 98.49 | 465 | 7 | 0 | 182380 | 5559 | 6023 | extract | 0.0 | 807 |
| 34 | 20 | 98.49 | 465 | 7 | 0 | 182815 | 19783 | 20247 | extract | 0.0 | 807 |
| 35 | 26 | 98.49 | 465 | 7 | 0 | 182880 | 18879 | 19343 | extract | 0.0 | 807 |
| 46 | 255 | 98.49 | 465 | 7 | 0 | 183179 | 4982 | 5446 | extract | 0.0 | 807 |
| 52 | 243 | 98.49 | 465 | 7 | 0 | 183381 | 4532 | 4996 | extract | 0.0 | 807 |
| 61 | 393 | 98.49 | 465 | 7 | 0 | 183648 | 7217 | 7681 | extract | 0.0 | 807 |
| 64 | 254 | 98.49 | 465 | 7 | 0 | 183818 | 4716 | 5180 | extract | 0.0 | 807 |
| 67 | S5611 | 98.49 | 465 | 7 | 0 | 184143 | 4804 | 5268 | extract | 0.0 | 807 |
| 82 | 11-004 | 98.49 | 465 | 7 | 0 | 184297 | 19518 | 19982 | extract | 0.0 | 807 |
| 84 | IAL2229 | 98.49 | 465 | 7 | 0 | 184635 | 4530 | 4994 | extract | 0.0 | 807 |
| 90 | CN100 | 98.49 | 465 | 7 | 0 | 184814 | 5350 | 5814 | extract | 0.0 | 807 |

Download FASTA | FASTA with flanking | Table (tab-delimited text) | Excel format

Clicking any of the 'extract' buttons will display the matched sequence along with a translated sequence and flanking sequences.

| | | | | | | |
|--------|---------|---------|-------------------------|---|-----|-----|
| 181867 | 35889 | 36353 | extract | → | 0.0 | 807 |
| 182004 | 36775 | 37239 | extract | → | 0.0 | 807 |
| 182318 | 19090 | 19554 | extract | → | 0.0 | 807 |
| 215673 | 4534 | 4998 | extract | ← | 0.0 | 807 |
| 8 | 1363524 | 1363988 | extract | → | 0.0 | 785 |
| 182380 | 5559 | 6023 | extract | → | 0.0 | 807 |
| 182815 | 19783 | 20247 | extract | → | 0.0 | 807 |

[Query](#) | [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
[Breakdown](#): [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
[Links](#): [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Extracted sequence: Seqbin id#:182318 (19090-19554)

sequence bin id#182318

| sequence | method | start | end | length | orientation | translation |
|----------|--|-------|-------|--------|-------------|-------------|
| Solexa | CGGACGGGT TGAACAGC ATCCGAATA CGCCGAAARA GCGCGGCCA AACTCGCCGA ATGTCCGGG TTGCGGTTTG TCAGCGGCC GAACAAATTT GAAGCCTTGG GCGGACGCGA TGCCGCCGTT GCGGCTTCGG | 19090 | 19554 | 465 | → | |

```

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 * F1
G T G L N S H P E Y A E K A A A K L A E L S G L P F V S A P N K F F2
A R V * T A I P N T P K K P P P N S P N C P A C R L S A R R T N L F3
1 CGGACGGGTTTGAACAGCATTCCGGAATACGCCGAAAGCGCCGCGCAACTCGCCGAATTGTCCGGCTTTCGCGTTTGTTCAGCGCGCGCAACAAATTT 100
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
S L G R T R C R R C R F G R I E N A G G K P E * N R Q R H P L A G F1
E A L G G R D A A V A A S G A L K T L A A S L N K I A N D I R W L A F2
K P W A D A M P P L P L R A H * K R W R Q A * I K S P T T S A G W F3
101 GAAGCCTTGGGCGGACGCGATCCGCGCTTGCCTTTCGGGCGCGATTGAAACGCTGGCGCAAGCCTGAATATAATGCGCAACGACATCCGCTGGCTGG 200
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
K R P A L R F G R N Q N P R K R A G F V H H A G Q S Q P D P M R S F1
S G P R C G L G E I K I P E N E P G S S I M P G K V N P T Q C E A F2
Q A A R A A V W A K S K S P K T S R V R P S C R A K S T R P N A K R F3
201 CAAGCGCCGCGCGCTTTCGGCGAATCAAAATCCCGGAAACGCGCGCGGTTTCGTCATCATGCGCGCAAGTCAACCGGACCGCAATTCGGAAGC 300
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
D D H G V L P S V R Q R R Y H R Y G G R V G Q F R A E R L Y A R H R F1
M T M V C C Q V F G N D V T I G M A G A S G N F E L N V Y M P V I F2
P W C A A K C S A T T L P S V W R A R R A I S S * T S I C P S S F3
301 GATGACCATGTTGTGTCGCAAGTGTTCGCGACGCGTTACCATCGGTATGCGCGCGCGCTCGCGCAATTTGCAAGTGAACGCTATATGCGCGTCATC 400
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
L Q P L A I H P P V G R R V Q Q L Q R T L R R R H * T R T G K N R F1
A Y N L L Q S I R L L G D A C N S F N E H C A V G I E P V P E K I D F2
P T T S C N P S A C W A T R A T A S T N T A P S A L N P Y R K K S F3
401 GCTTACACCTCTTTCGCAATCCATCCGCTGTTGGGCGACGCGTGCACGAGCTTCAACGACACTGCGCGCTGCGCATGGAACCGTACCGGAAATAATCG 500

```

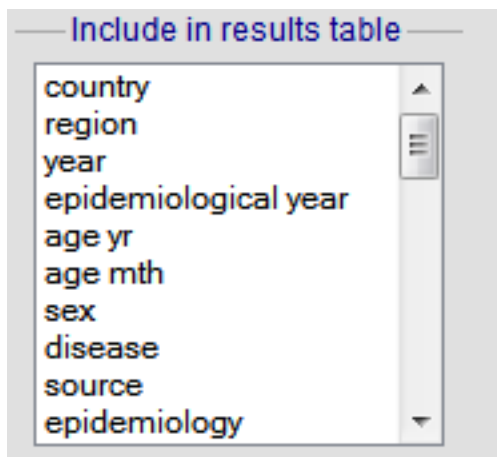
At the bottom of the results table are links to export the matching sequences in FASTA format, (optionall) including flanking sequences. You can also export the table in tab-delimited text or Excel formats.

| | | | | | | | | | | | | |
|----|---------|-------|-----|---|---|--------|-------|-------|-------------------------|---|-----|-----|
| 82 | 11-004 | 98.49 | 465 | 7 | 0 | 184297 | 19518 | 19982 | extract | → | 0.0 | 807 |
| 84 | IAL2229 | 98.49 | 465 | 7 | 0 | 184635 | 4530 | 4994 | extract | ← | 0.0 | 807 |
| 90 | CN100 | 98.49 | 465 | 7 | 0 | 184814 | 5350 | 5814 | extract | ← | 0.0 | 807 |

[Download FASTA](#) | [FASTA with flanking](#) | [Table \(tab-delimited text\)](#) | [Excel format](#)

12.7.1 Include in results table fieldset

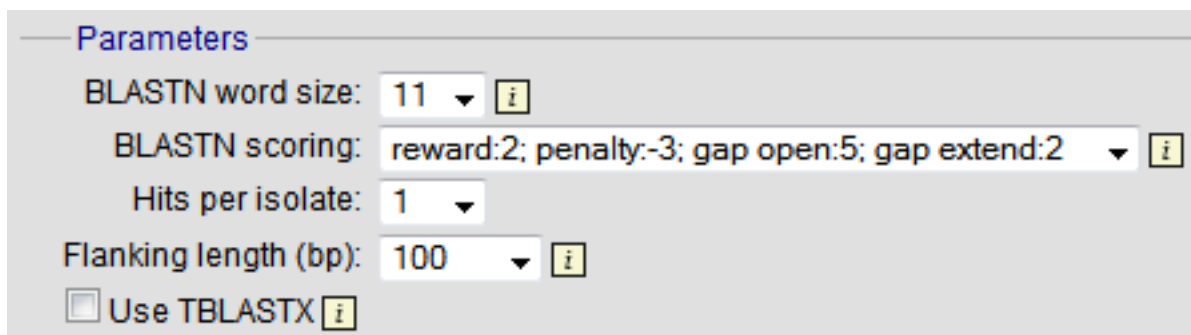
This selection box allows you to choose which isolate provenance fields will be included in the results table.



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

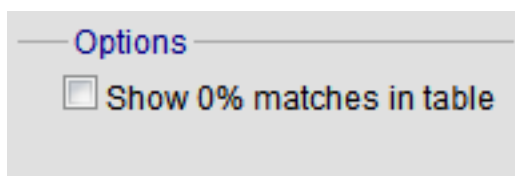
12.7.2 Parameters fieldset

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



- BLASTN word size - This is the length of the initial identical match that BLAST requires before extending a match (default: 11). Increasing this value improves speed at the expense of sensitivity.
- BLASTN scoring - This is a dropdown box of combinations of identical base rewards; mismatch penalties; and gap open and extension penalties. BLASTN has a constrained list of allowed values which reflects the available options in the list.
- Hits per isolate - By default, only the best match is shown. Increase this value to the number of hits you'd like to see per isolate.
- Flanking length - Set the size of the upstream and downstream flanking sequences that you'd like to include.
- Use TBLASTX - This compares the six-frame translation of your nucleotide query sequence against the six-frame translation of the contig sequences. This is significantly slower than using BLASTN.

12.7.3 No matches





Click this option to create a row in the table indicating that a match was not found. This can be useful when screening a large number of isolates.


12.7.4 Filter fieldset

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

Restrict included sequences by

Sequence method: 

Project: 

Experiment: 

Available options are:

- Sequence method - Choose to only analyse contigs that have been generated using a particular method. This depends on the method being set when the contigs were uploaded.
- Project - Only include isolates belonging to the chosen project. This enables you to select all isolates and filter to a project.
- Experiment - Contig files can belong to an experiment. How this is used can vary between databases, but this enables you to only include contigs from a particular experiment.

12.8 BURST

BURST is an algorithm used to group MLST-type data based on a count of the number of profiles that match each other at specified numbers of loci. The analysis is available for both sequence definition database and isolate database schemes that have primary key fields set. The algorithm has to be *specifically enabled* by an administrator. Analysis is limited to 1000 or fewer records.

The plugin can be accessed following a query by clicking the 'BURST' button in the Analysis list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

| | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------------------------------|------------------------------|------------------------|------|---------------|----------------------------|------|------|------|
| 6 | M00282207 | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 0021/84 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown:

Analysis:

Export:

Page:

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

PubMLST

[Query: Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
[Breakdown: Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
[Links: Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

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:

- Divide strains into groups according to their allelic profiles.
- Count the number of Single Locus Variants (SLV), Double Locus Variants (DLV) and Satellites (SAT) for each sequence type (ST).
- Identify the potential Ancestral Type (AT). These are shown with an asterisk next to their names in the results table.

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

Select scheme: MLST

Group definition: profiles match at: n-2 | loci to any other member of the group [n = number of loci in scheme].

☒ Shade variant rings
☐ Hide variant names (useful for overview if names start to overlap)

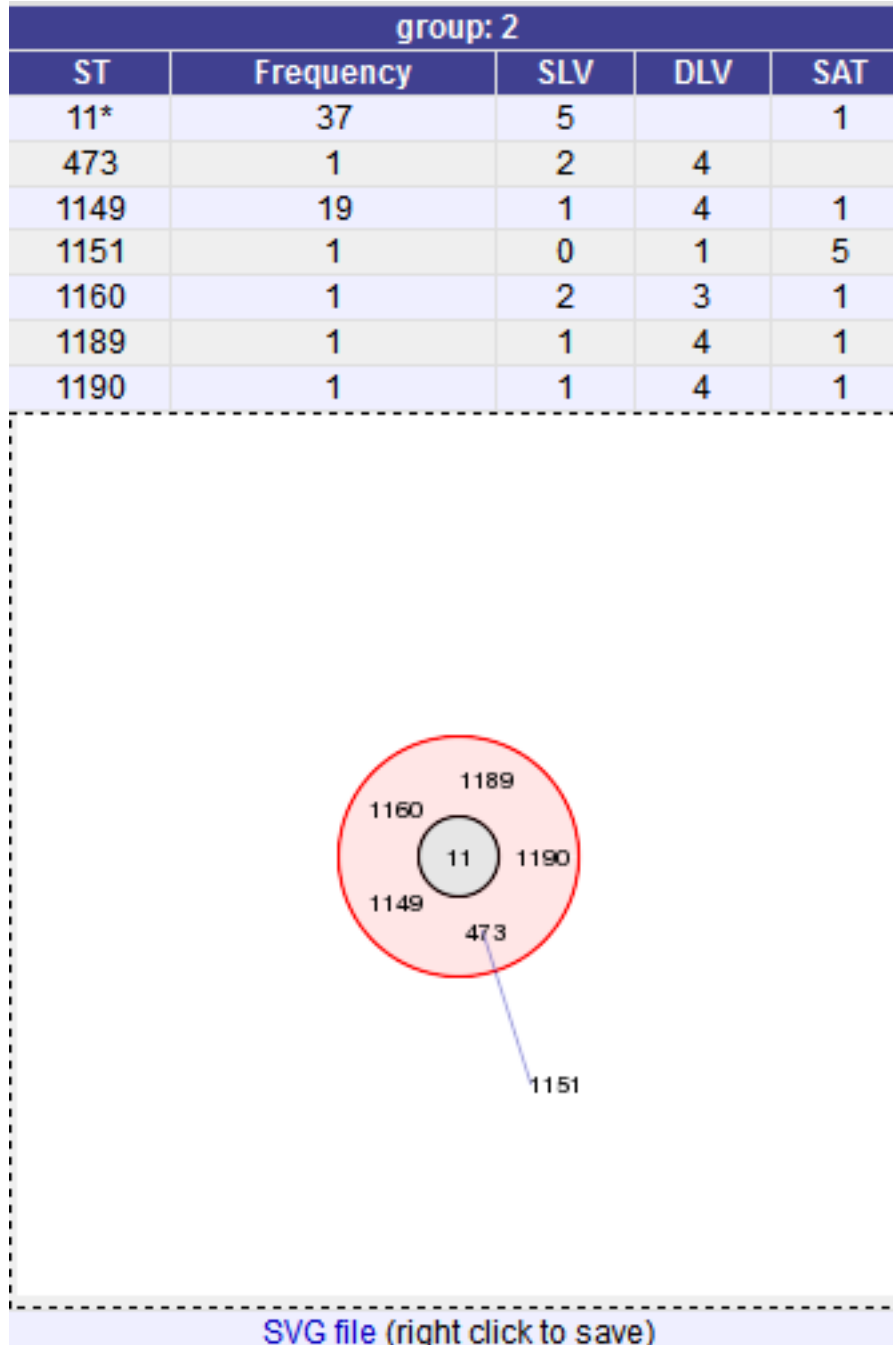
Action

Submit

Modifying the group definition affects the size of groups and how they link together. By default, the definition is n-2 (where n is the number of loci), so for example on a 7 locus MLST scheme groups contain STs that match at 5 or more loci to any other member of the group.

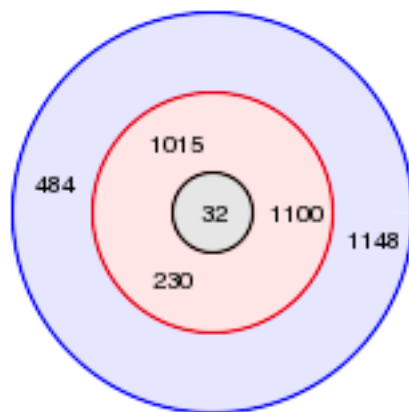
Click Submit.

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



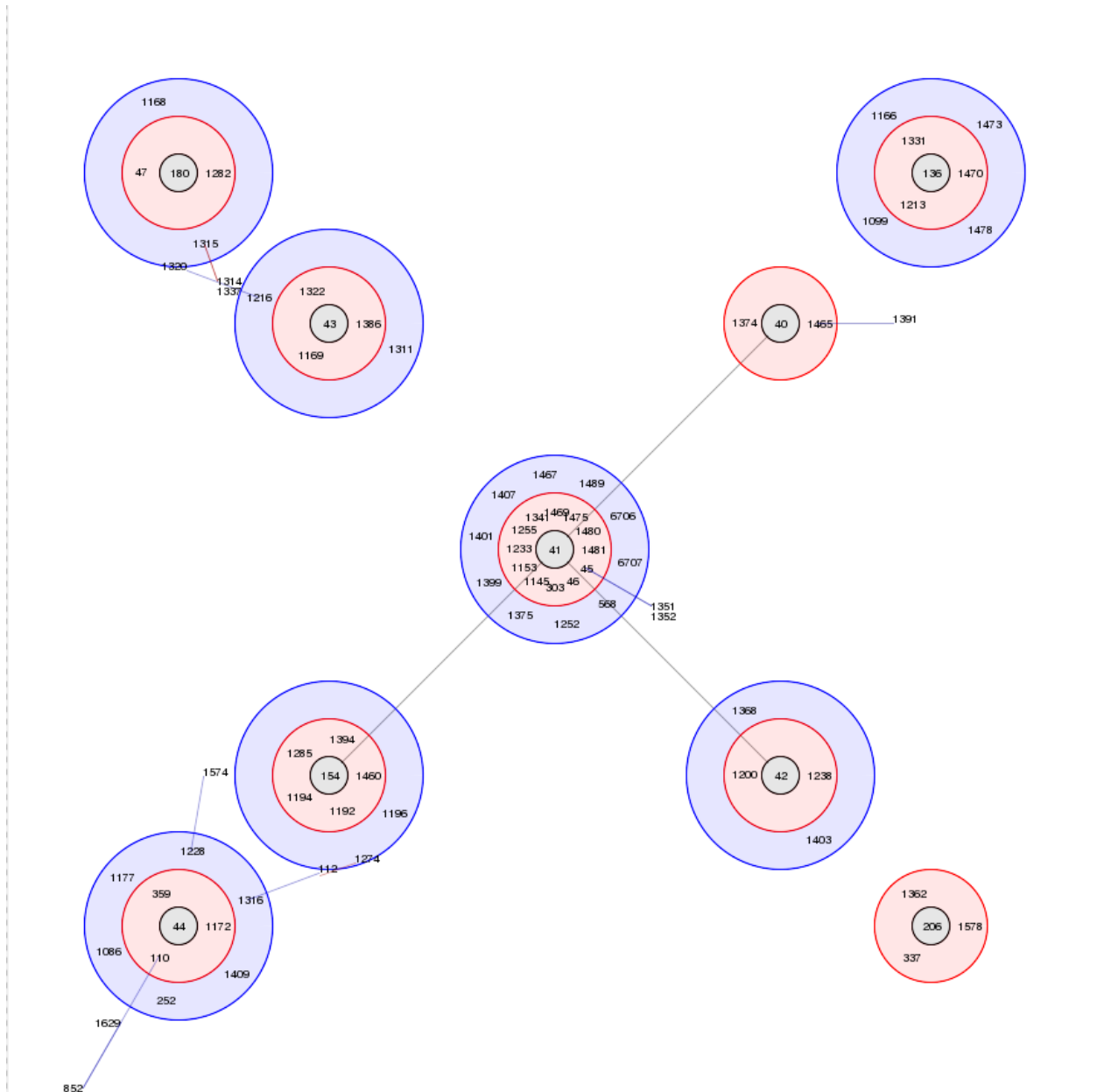
SLV profiles that match the central profile are shown within a red circle surrounding the central profile. Most distant profiles (triple locus variants) may be linked with a line. Larger groups may additionally have DLV profiles. These are shown in a blue circle.

| group: 6 | | | | |
|----------|-----------|-----|-----|-----|
| ST | Frequency | SLV | DLV | SAT |
| 32* | 2 | 3 | 2 | |
| 230 | 1 | 1 | 3 | 1 |
| 484 | 1 | 0 | 3 | 2 |
| 1015 | 1 | 1 | 4 | |
| 1100 | 1 | 1 | 2 | 2 |
| 1148 | 1 | 0 | 4 | 1 |



[SVG file \(right click to save\)](#)

Groups can get very large, where linked profiles form sub-groups and an attempt is made to depict these.



12.9 Codon usage

The codon usage plugin for isolate databases calculates the absolute and relative synonymous codon usage by isolate and by locus.

The function can be selected by clicking the 'Codon usage' link in the Analysis section of the main contents page.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

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.

| | | | | | | | | | | | |
|----|-----------|-------------|----------------|------------------------------|------------------------------|----|------|----------------------------|------|------|------|
| 5 | M00240227 | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 6 | M00282207 | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 002184 | | Czech Republic | 1984 | invasive (unspecified/other) | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown: [Fields](#) | [Two Field](#) | [Polymorphic sites](#) | [Combinations](#) | [Schemes/alleles](#) | [Publications](#) | [Sequence bin](#) | [Tag status](#)

Analysis: [BURST](#) | [Codons](#) | [Presence/Absence](#) | [Genome Comparator](#) | [BLAST](#)

Export: [Dataset](#) | [Contigs](#) | [Sequences](#)

Page: [1](#) | [2](#) | [3](#) | [4](#) | [Last](#)

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.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Codon usage analysis Toggle: [?]

This plugin will analyse the codon usage for individual loci and overall for an isolate. Only loci that have a corresponding database containing sequences, or with sequences tagged, can be included. It is important to note that correct identification of codons can only be achieved for loci for which the correct ORF has been set (if they are not in reading frame 1). Partial sequences from the sequence bin will not be analysed. Please check the loci that you would like to include. Output is limited to 500 records.

Select ids
Paste in list of ids to include, start a new line for each. Leave blank to include all ids.

1
2
3
4
5
6

Include in identifier

isolate
country
region
year
epidemiological year
age yr
age mth
sex
disease
source

Loci

16S_rDNA
16S_rRNA (SSU_rRNA)
_23S_rRNA
abcZ
abcZ (NEIS1015)
aceF (NEIS1279)
ackA2 (NEIS1727)
ackA (NEIS1729)

All None Paste list

Schemes

Genetic Information Process
Metabolism
Pilin
Typing
MLST
Finetyping antigens
16S
Antigen genes

Sequence retrieval
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 (defined alleles will still be used)

Codons
Select codon order:
☒ alphabetical
☐ C or G ending codons first

Action
Submit

Click submit. The job will be submitted to the queue and will start running shortly. Click the link to follow the job progress and view the output.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Codon usage analysis Toggle: [?]

This analysis has been submitted to the job queue.
Please be aware that this job may take some time depending on the number of sequences to analyse and how busy the server is.

[Follow the progress of this job and view the output.](#)

Four tab-delimited text files will be created.

- Absolute frequency of codon usage by isolate
- Absolute frequency of codon usage by locus
- Relative synonymous codon usage by isolate
- Relative synonymous codon usage by locus

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Job status viewer

Status

| | |
|--------------|-------------------------------|
| Job id: | BIGSdb_13269_1405586315_76138 |
| Submit time: | 2014-07-17 09:38:35 |
| Status: | finished |
| Start time: | 2014-07-17 09:38:39 |
| Progress: | 100% |
| Stop time: | 2014-07-17 09:39:05 |
| Total time: | 26 seconds |

Output

- Absolute frequency of codon usage by isolate
- Absolute frequency of codon usage by locus
- Relative synonymous codon usage (RSCU) by isolate
- Relative synonymous codon usage (RSCU) by locus
- Tar file containing output files

Please note that job results will remain on the server for 7 days.

12.10 Unique combinations

The Unique Combinations plugin calculates the frequencies of unique file combinations within an isolate dataset. Provenance fields, composite fields, allele designations and scheme fields can be combined.

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

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations**
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

Alternatively, it can be accessed following a query by clicking the 'Combinations' button in the Breakdown list at the bottom of the results table. This will run the analysis on the dataset returned from the query. Please note that the list of functions here may vary depending on the setup of the database.

| | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------|------------------------------|------------------------|----|------|----------------------------|------|------|------|
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 0021/84 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown:

Analysis:

Export:

Page:

Select the combination of fields to analyse, e.g. serogroup and finotyping antigens.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [\[i\]](#)

Frequencies of field combinations

Here you can determine the frequencies of unique field combinations in the dataset. Please select your combination of fields. Select loci either from the locus list or by selecting one or more schemes to include all loci (and/or fields) from a scheme.

Isolate fields

☐ id ☒ serogroup ☐ rifampicin ☐ strain designation

☐ isolate ☐ MLEE designation ☐ rifampicin range ☐ All ☐ None

☐ aliases ☐ serotype ☐ ciprofloxacin ☐ ciprofloxacin range

☐ country ☐ sero subtype ☐ pending assembly ☐ assembly status

☐ continent ☐ ET no ☐ ENA accession ☐ private project

☐ region ☐ penicillin ☐ comments ☐ sender

☐ year ☐ penicillin range ☐ date entered ☐ curator

☐ epidemiological year ☐ amoxicillin ☐ datestamp

☐ age yr ☐ sulphonamide

☐ age mth ☐ ceftriaxone ☐ ceftriaxone range

☐ sex ☐ chloramphenicol ☐ cefotaxime

☐ disease ☐ source ☐ cefotaxime range

☐ epidemiology ☐ species

Composite fields ☐ strain designation

Loci


16S_rDNA
16S_rRNA (SSU_rRNA)
23S_rRNA
abcZ
abcZ (NEIS1015)
aceF (NEIS1279)
ackA2 (NEIS1727)
acnA (NEIS1729)

Schemes

Metabolism
Pilin
Typing
MLST
Finotyping antigens
16S
Antigen genes

☒ Include all fields from selected schemes
☒ Include all loci from selected schemes

Click submit. When the analysis has completed you will see a table showing the unique combinations of the selected fields along with the frequency and percentage of the combination.


[Query: Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
[Breakdown: Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
[Links: Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [?]

Frequencies of field combinations

Number of unique combinations: 2939

The percentages may add up to more than 100% if you have selected loci or scheme fields with multiple values for an isolate.

| serogroup | PorA VR1 | PorA VR2 | FetA VR | Frequency | Percentage |
|-----------|----------|----------|---------|-----------|------------|
| B | 7-2 | 4 | F1-5 | 625 | 5.65 |
| W | 5 | 2 | F1-1 | 513 | 4.64 |
| B | 19 | 15 | F5-1 | 385 | 3.48 |
| C | 5 | 2 | F3-6 | 283 | 2.56 |
| - | 18-10 | 43 | F3-14 | 281 | 2.54 |
| B | 7 | 16 | F3-3 | 252 | 2.28 |
| C | 5-1 | 10-8 | F3-6 | 222 | 2.01 |
| B | 22 | 9 | F5-12 | 171 | 1.55 |
| B | 22 | 14 | F5-5 | 167 | 1.51 |
| A | 5-2 | 10 | F3-5 | 135 | 1.22 |
| B | 19-1 | 15-11 | F5-1 | 107 | 0.97 |
| Y | 5-1 | 10-1 | F4-1 | 96 | 0.87 |
| - | 5 | 2 | F1-1 | 93 | 0.84 |
| Y | 5-2 | 10-1 | F4-1 | 79 | 0.71 |
| C | 5 | 2 | F5-8 | 79 | 0.71 |
| W | 18-1 | 3 | F4-1 | 77 | 0.70 |
| B | 22-1 | 14 | F4-1 | 77 | 0.70 |
| C | 5 | 2 | F3-3 | 76 | 0.69 |
| B | 18-1 | 3 | F1-5 | 72 | 0.65 |
| Y | 5-1 | 2-2 | F5-8 | 69 | 0.62 |
| B | 19 | 15 | F1-14 | 61 | 0.55 |
| B | 7-2 | 13-2 | F1-5 | 61 | 0.55 |
| B | 7-2 | 16 | F3-3 | 56 | 0.51 |

The table can be downloaded in tab-delimited text or Excel formats by clicking the links at the bottom of the page.

| | | | | | |
|----|------|--------|--------|---|------|
| C | 22 | 14-6 | F1-7 | 1 | 0.01 |
| B | 19 | 13-2 | F5-1 | 1 | 0.01 |
| B | 12-1 | 13-1 | F3-29 | 1 | 0.01 |
| NG | 7-2 | 16-126 | F3-3 | 1 | 0.01 |
| B | 7-2 | 30 | F5-1 | 1 | 0.01 |
| B | 7 | 30-3 | F1-14 | 1 | 0.01 |
| B | 5-1 | 10-10 | F3-16 | 1 | 0.01 |
| - | 7-2 | 30-3 | F1-7 | 1 | 0.01 |
| C | 5-2 | 10 | F1-7 | 1 | 0.01 |
| NG | 18-1 | 3 | F1-34 | 1 | 0.01 |
| NG | 22 | 14-6 | F4-2 | 1 | 0.01 |
| B | 7-2 | 4 | F1-88 | 1 | 0.01 |
| NG | 5-1 | 10-62 | F1-3 | 1 | 0.01 |
| Y | 12-3 | 4 | F4-1 | 1 | 0.01 |
| NG | 7-2 | 16 | F1-104 | 1 | 0.01 |
| X | 12-1 | 16-52 | F3-9 | 1 | 0.01 |
| NG | 18-4 | 35-34 | F4-1 | 1 | 0.01 |

[Download as tab-delimited text](#)
[Download in Excel format](#)

12.11 Polymorphisms

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

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

| | | | | | | | | | | | | |
|----|---------|-------------|----------------|------|------------------------------|------------------------|----|------|----------------------------|------|------|------|
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-19 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown: Fields Two Field **Polymorphic sites** Combinations Schemes/alleles Publications Sequence bin Tag status

Analysis: BURST Codons Presence/Absence Genome Comparator BLAST

Export: Dataset Contigs Sequences

Page: 1 2 3 4 5 6 7 8 > Last

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

Query: Search | Browse | Profile/ST | List
 Breakdown: Isolate fields | Scheme/alleles | Publications
 Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Toggle: [i]

Polymorphic site analysis

This tool will analyse the polymorphic sites in the selected locus for the current isolate dataset.

If more than 50 sequences have been selected, the job will be run by the offline job manager which may take a few minutes (or longer depending on the queue). This is because sequences may have gaps in them and consequently need to be aligned which is a processor- and memory- intensive operation.

Loci

- *16S_rDNA
- 16S_rRNA (SSU_rRNA)
- 23S_rRNA
- abcZ**
- abcZ (NEIS1015)
- aceF (NEIS1279)
- ackA2 (NEIS1727)
- acnA (NEIS1729)

Options

If both allele designations and tagged sequences exist for a locus, choose how you want these handled: [1]

☒ Use allele sequence retrieved from external database

☐ Use sequences tagged from the bin

☒ Analyse single example of each unique sequence

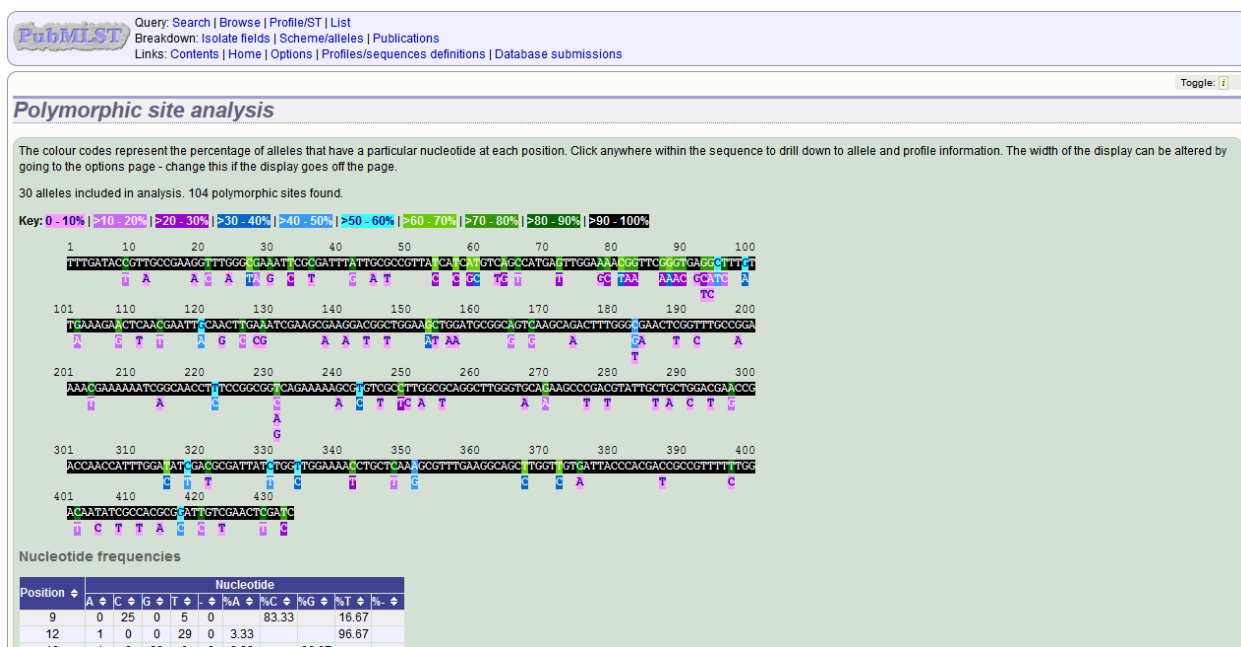
☒ Exclude incomplete sequences

Action

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.



12.12 Presence/absence

This plugin displays the status of loci for isolate records. It will show whether a locus has been designated with an allele name, has a sequence tag, or both.

The function can be selected by clicking the 'Presence/absence status of loci' link in the 'Analysis' section of the main contents page.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34221
- Last updated: 2015-07-02
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs
- Sequences - XMFA / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci**
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

Alternatively, it can be accessed following a query by clicking the 'Presence/Absence' button in the Analysis list at the bottom of the results table. Please note that the list of functions here may vary depending on the setup of the database.

| | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------|------------------------------|------------------------|----|------|----------------------------|------|------|------|
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 002184 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown:

Analysis:

Export:

Page:

Enter the ids of the isolate records to analyse - these will be already entered if you accessed the plugin following a query. Select the loci you would like to analyse, either from the dropdown loci list, and/or by selecting one or more schemes.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Export presence/absence status of loci - *Neisseria* PubMLST

This script will export data showing whether a locus has had an allele designated, a sequence tagged, or both. Please check the loci that you would like to include. Alternatively select one or more schemes to include all loci that are members of the scheme.

Select ids: Paste in list of ids to include, start a new line for each. Leave blank to include all ids.

1
2
3
4
5
6

Include in identifier: isolate, country, region, year, epidemiological year, age yr, age mth, sex, disease, source

Loci: 16S_rDNA, 16S_rRNA (SSU_rRNA), 23S_rRNA, abcZ, abcZ (NEIS1015), aceF (NEIS1279), ackA2 (NEIS1727), acnA (NEIS1729)

Schemes: Genetic information processing, Metabolism, Pilin, Typing, MLST, Finotyping antigens, 16S, Antigen genes

Options: Mark present if: ☒ either designations or tags set, ☐ allele designations defined, ☐ sequence tags defined. Symbol for present: O, Symbol for absent: X, ☐ Generate distance matrix

Action:

Click submit. The job will be submitted to the queue and will start running shortly. Click the link to follow the job progress and view the output.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Export presence/absence status of loci - *Neisseria* PubMLST

This analysis has been submitted to the job queue.

[Follow the progress of this job and view the output.](#)

When complete, a single text file will have been generated.



Job status viewer

Status

Job id: BIGSdb_31442_1405591304_69061
 Submit time: 2014-07-17 11:01:44
 Status: finished
 Start time: 2014-07-17 11:01:44
 Progress: 100%
 Stop time: 2014-07-17 11:01:49
 Total time: 4 seconds

Output

- Main output file

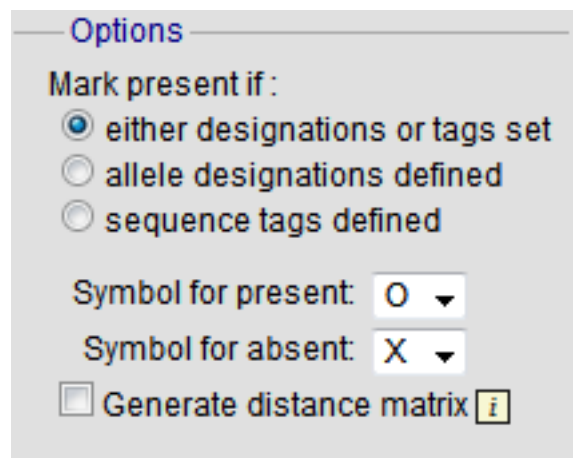
Please note that job results will remain on the server for 7 days.

This is a tab-delimited text file that uses 'O' to represent presence and 'X' to represent a missing locus designation or tag.

| id | pgm | adk | abcZ | pdhC | gdh | fumC | aroE |
|----|-----|-----|------|------|-----|------|------|
| 1 | O | O | O | O | O | O | O |
| 2 | O | O | O | O | O | O | O |
| 3 | O | O | O | O | O | O | O |
| 4 | O | O | O | O | O | O | O |
| 5 | O | O | O | O | O | O | O |
| 6 | O | O | O | O | O | O | O |
| 7 | O | O | O | O | O | O | O |
| 8 | O | O | O | O | O | O | O |
| 9 | O | O | O | O | O | O | O |
| 10 | O | O | O | O | O | O | O |

12.12.1 Options

There are a number of options that can be selected to modify the output.




Options

Mark present if :

☒ either designations or tags set
☐ allele designations defined
☐ sequence tags defined

Symbol for present: O ▼

Symbol for absent: X ▼

☐ Generate distance matrix 

With these you can change the symbols used and whether designations, or tags, or both are counted.

You can also choose to generate a distance matrix based on presence/absence.

12.13 Tag status

The tag status plugin displays a graphical representation of the status of loci designations or tags for isolate data. It is accessed following a query by clicking the ‘Tag status’ button in the Breakdown section at the bottom of the results table.

The screenshot shows the BIGSdb interface with the following components:

- Isolate provenance/phenotype fields:** A dropdown menu for 'id' and a search box with '10' entered.
- Display/sort options:** A dropdown menu for 'Order by: id' and a dropdown menu for 'Display: 25 records per page'.
- Action:** 'Reset' and 'Submit' buttons.
- 9 records returned. Click the hyperlinks for detailed information.**
- Table:** A table with columns: id, isolate, aliases, country, year, disease, species, serogroup, ST, MLST (clonal complex), and Finotyping antigens (PorA VR1, PorA VR2, FetA VR).
- Analysis tools:** A section with buttons for 'Breakdown', 'Analysis', and 'Export'. The 'Breakdown' section includes 'Fields', 'Two Field', 'Polymorphic sites', 'Combinations', 'Schemes/alleles', 'Publications', 'Sequence bin', and 'Tag status' (highlighted with a red box).

Select the loci you would like to analyse.

The screenshot shows the 'Tag status' page with the following components:

- Query:** Search | Browse | Profile/ST | List
- Breakdown:** Isolate fields | Scheme/alleles | Publications
- Links:** Contents | Home | Options | Profiles/sequences definitions | Database submissions
- Tag status:** A section with a tree view of loci. The tree view shows a hierarchy of loci, with 'All loci' selected and highlighted with a red box.

You should see a series of bars representing loci. The colour of these bars designates whether they have an allele designation only, a sequence tag only, both designations or tags, or whether they have flags set.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Tag status

Select schemes or groups of schemes within the tree. A breakdown of the individual loci belonging to these schemes will then be performed.

Bars represent loci by schemes arranged in alphabetical order. If a locus appears in more than one scheme it will appear more than once in this graphic. Click on the id hyperlink for a detailed breakdown for an isolate.

Key
Allele designated only | Sequence tagged only | Allele designated + sequence tagged | Flagged

2463 loci selected:

| Id | Isolate |
|----|-----------|
| 1 | A4/M1027 |
| 2 | 120M |
| 3 | M00242905 |
| 4 | M1027 |
| 5 | M00240227 |
| 6 | M00282207 |
| 7 | 7891 |
| 8 | M00242007 |
| 9 | 0021/84 |

Hovering the mouse over the bars will indicate the scheme represented.

Note: Loci will be represented more than once if they are members of multiple selected schemes.

Clicking any of the isolate id hyperlinks navigates to a page that breaks down the exacty status for all loci of that isolate.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Tag status: Isolate id#1 (A4/M1027)

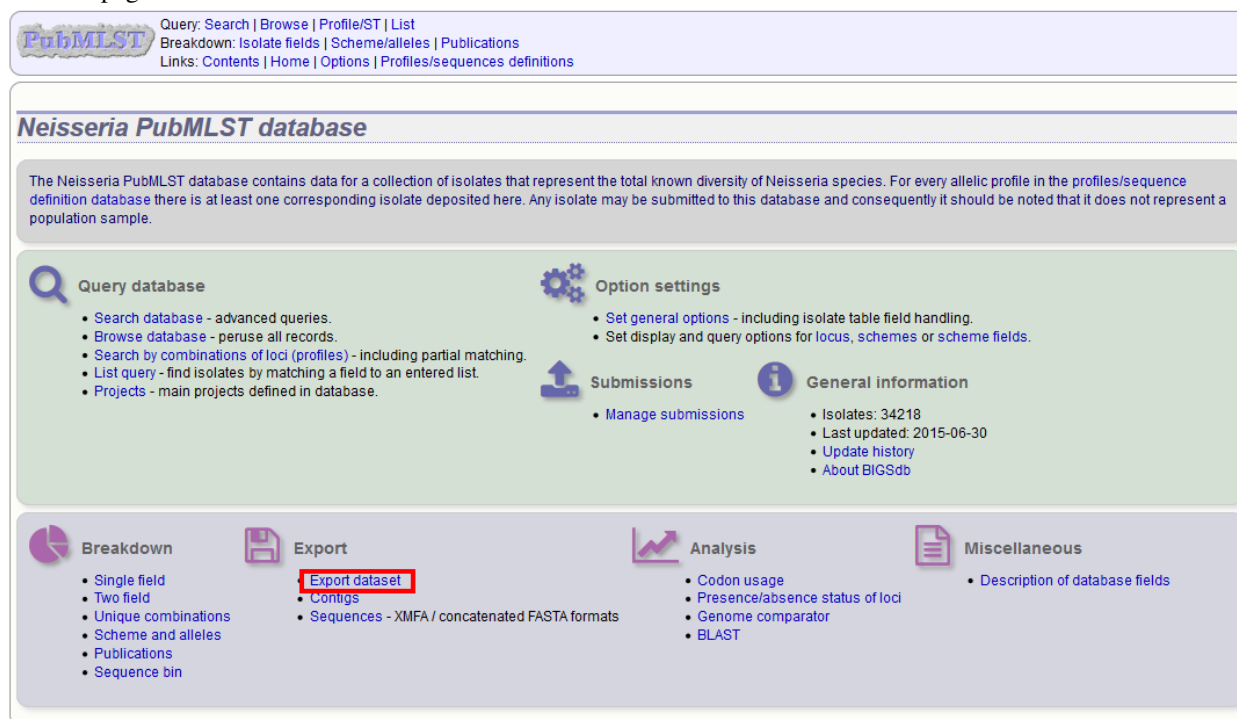
| Scheme | Locus | Allele designation(s) | Sequence tag |
|-----------------------------|---------------------|-----------------------|---------------------|
| MLST | abcZ | 1 | |
| | adk | 3 | |
| | aroE | 3 | |
| | fumC | 1 | |
| | gdh | 4 | |
| | pdhC | 2 | |
| Finotyping antigens | pgm | 3 | |
| | PorA VR1 | 5-2 | |
| 16S | PorA VR2 | 10 | |
| | FeA VR | F1-5 | |
| ADP-heptose biosynthesis | SSU rRNA (16S rRNA) | 45 | |
| | NEIS0769 (hldA) | 2 | |
| | NEIS0773 (hldD) | 2 | |
| | NEIS2014 (gmhB) | 2 | |
| | NEIS2055 (hldC) | 2 | |
| Aminoacyl-tRNA biosynthesis | NEIS2070 (gmhA) | | |
| | NEIS0007 | 2 | |
| | NEIS0164 | 2 | |
| | NEIS0326 | 61 | |
| | NEIS0381 | 2 | |
| | NEIS0383 | 2 | |
| | NEIS0672 | | internal stop codon |
| | NEIS0676 | 15 | |
| | NEIS0681 | 5 | |
| | NEIS0794 | 5 | |
| | NEIS1277 | 3 | |
| | NEIS1290 (gatC) | 3 | |
| | NEIS1291 | 13 | |
| | NEIS1293 | 9 | |
| | NEIS1361 | 3 | |
| NEIS1408 | 71 | | |
| NEIS1436 | 5 | | |
| NEIS1478 | 4 | | |
| NEIS1518 | 3 | | |
| NEIS1602 | 3 | | |

There is a column each for allele designations and sequence tags. If an allele designation is defined, the allele identifier is displayed. Cells shaded in blue show that the designation or tag is present, whereas red indicates that they are absent.

Data export plugins

13.1 Isolate record export


You can export the entire isolate recordset by clicking the ‘Export dataset’ link in the Export section of the main contents page.




PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database


The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the [profiles/sequence definition database](#) there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.


Query database


- [Search database](#) - advanced queries.
- [Browse database](#) - peruse all records.
- [Search by combinations of loci \(profiles\)](#) - including partial matching.
- [List query](#) - find isolates by matching a field to an entered list.
- [Projects](#) - main projects defined in database.


Option settings


- [Set general options](#) - including isolate table field handling.
- [Set display and query options for locus, schemes or scheme fields.](#)


Submissions


- [Manage submissions](#)


General information


- Isolates: 34218
- Last updated: 2015-06-30
- [Update history](#)
- [About BIGSdb](#)


Breakdown


- [Single field](#)
- [Two field](#)
- [Unique combinations](#)
- [Scheme and alleles](#)
- [Publications](#)
- [Sequence bin](#)


Export

- [Export dataset](#)
- [Contigs](#)
- [Sequences](#) - XMFA / concatenated FASTA formats


Analysis

- [Codon usage](#)
- [Presence/absence status of loci](#)
- [Genome comparator](#)
- [BLAST](#)


Miscellaneous

- [Description of database fields](#)

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.

99 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 > Last

| Isolate fields | | | | | | | ML ST | | Finotyping antigens | | | |
|----------------|-----------|-------------|----------------|------|------------------------------|------------------------|-----------|------|----------------------------|----------|----------|---------|
| id | isolate | aliases | country | year | disease | species | serogroup | ST | clonal complex | PorA VR1 | PorA VR2 | FetA VR |
| 1 | A4M1027 | B1; Z1001 | USA | 1937 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 5-2 | 10 | F1-5 |
| 2 | 120M | B35; Z1035 | Pakistan | 1967 | meningitis and septicaemia | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 3 | M00242905 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1099 | | 19 | 15 | |
| 4 | M1027 | B43; Z1043 | USA | 1937 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | | | |
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 002184 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown: Fields Two Field Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status

Analysis: BURST Codons Presence/Absence Genome Comparator BLAST

Export: Dataset Contigs Sequences

Page: 1 2 3 4 > Last

Select the isolate fields and schemes to include.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Toggle: [?](#)

Export dataset

This script will export the dataset in tab-delimited text, suitable for importing into a spreadsheet. 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.

Isolate fields

- ☒ id
- ☒ isolate
- ☒ aliases
- ☒ country
- ☐ continent
- ☐ region
- ☐ year
- ☐ epidemiological year
- ☐ age yr
- ☐ age mth
- ☐ sex
- ☐ disease
- ☐ source
- ☐ epidemiology
- ☐ species
- ☐ serogroup
- ☐ MLEE designation
- ☐ serotype
- ☐ sero subtype
- ☐ penicillin
- ☐ penicillin range
- ☐ amoxicillin
- ☐ sulphonamide
- ☐ ceftriaxone
- ☐ ceftriaxone range
- ☐ cefotaxime
- ☐ cefotaxime range
- ☐ rifampicin
- ☐ rifampicin range
- ☐ ciprofloxacin
- ☐ ciprofloxacin range
- ☐ ET no
- ☐ pending assembly
- ☐ assembly status
- ☐ ENA accession
- ☐ private project
- ☐ comments
- ☐ sender
- ☐ curator
- ☐ date entered
- ☐ datestamp

Composite fields

- ☐ strain designation

All None

References

- ☐ references
- ☒ PubMed id
- ☐ Full citation

Loci

- *16S_rDNA
- 16S_rRNA (SSU_rRNA)
- _23S_rRNA
- abcZ (NEIS1015)
- aceF (NEIS1279)
- ackA2 (NEIS1727)
- acnA (NEIS1729)

All None

Schemes

- Genetic information processing
- Metabolism
- Pilin
- Typing
- MLST
- Finotyping antigens
- 16S
- Antigen genes

☒ Include all fields from selected schemes
☒ Include all loci from selected schemes

Options

- ☐ Include locus common names
- ☒ Export allele numbers
- ☐ Use one row per field
- ☐ Include isolate field in row (used only with 'one row' option)
- ☐ Export full allele designation record (used only with 'one row' option)

Molecular weights

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

Action

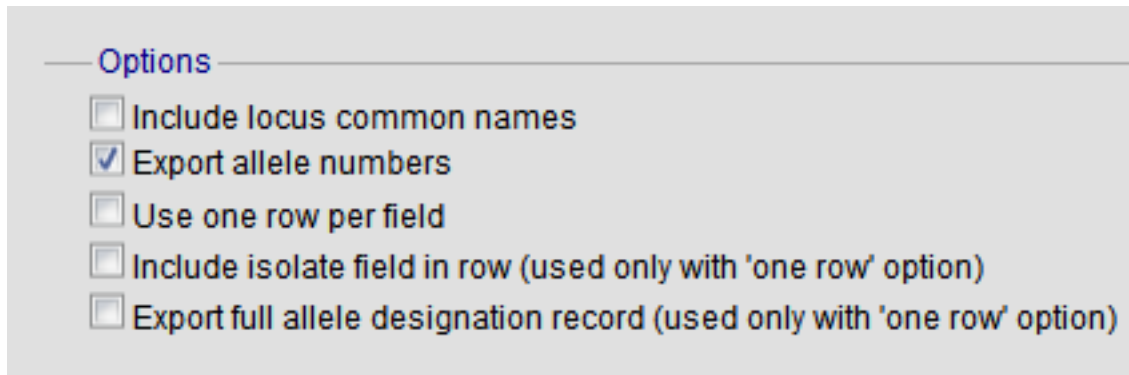
[Submit](#)

Click Submit.

You can then download the data in tab-delimited text or Excel formats.



13.1.1 Advanced options

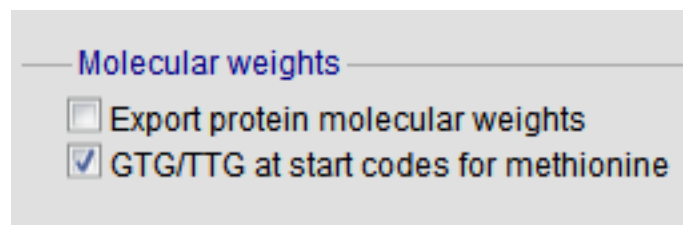


The options fieldset has the following options.

- Include locus common names - any common name for the locus is displayed in parentheses following the primary name.
- Export allele numbers - the allele designation is included for any locus included.
- Use one row per field - this is an alternative output format where instead of each locus and field having a separate column, each field is export on a separate row.
- Include isolate field in row - the name of the isolate is included as a separate column when exporting in 'one row per field' fomrmat.
- Export full allele designation record - export sender, curator and datestamp information as separate rows when exporting allele designation data.

13.1.2 Molecular weight calculation

The plugin can also calculate the predicted molecular weight of the gene product of any allele designated in the dataset.



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

13.2 Sequence export

You can export the sequences for any set of loci designated in isolate records, or belonging to scheme profiles in the sequence definition database.

The sequence export function can be accessed by clicking the ‘Sequences’ link in the Export section of the contents page.

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#)

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the [profiles/sequence definition database](#) there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- [Search database](#) - advanced queries.
- [Browse database](#) - peruse all records.
- [Search by combinations of loci \(profiles\)](#) - including partial matching.
- [List query](#) - find isolates by matching a field to an entered list.
- [Projects](#) - main projects defined in database.

Option settings

- [Set general options](#) - including isolate table field handling.
- Set display and query options for [locus](#), [schemes](#) or [scheme fields](#).

Submissions

- [Manage submissions](#)

General information

- Isolates: 34218
- Last updated: 2015-06-30
- [Update history](#)
- [About BIGSdb](#)

Breakdown

- [Single field](#)
- [Two field](#)
- [Unique combinations](#)
- [Scheme and alleles](#)
- [Publications](#)
- [Sequence bin](#)

Export

- [Export dataset](#)
- [Contigs](#)
- **Sequences** - XMFA / concatenated FASTA formats

Analysis

- [Codon usage](#)
- [Presence/absence status of loci](#)
- [Genome comparator](#)
- [BLAST](#)

Miscellaneous

- [Description of database fields](#)

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.

| | | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------|------------------------------|------------------------|----|------|----------------------------|------|------|------|--|
| 4 | M1027 | B43, Z1043 | USA | 1937 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | | | | |
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | | |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | | |
| 7 | 7991 | B54, Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 | |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | | |
| 9 | 002184 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | | |
| 10 | 6748 | B73, Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 18-1 | 3 | F5-1 | |
| 11 | 129 | B92, Z1092 | Germany | 1984 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F3-6 | |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | | |
| 13 | 139M | B99, Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup III | 5-2 | 10 | F5-1 | |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | | |
| 19 | S3131 | B213, Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 | |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | | |
| 24 | S4355 | B227, Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 | |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | | |

Analysis tools:

Breakdown: [Fields](#) [Two Field](#) [Polymorphic sites](#) [Combinations](#) [Schemes/alleles](#) [Publications](#) [Sequence bin](#) [Tag status](#)

Analysis: [BURST](#) [Codons](#) [Presence/Absence](#) [Genome Comparator](#) [BLAST](#)

Export: [Dataset](#) [Contigs](#) [Sequences](#)

Page: [1](#) [2](#) [3](#) [4](#) [Last](#)

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.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Export allele sequences in XMFA/concatenated FASTA formats - Neisseria PubMLST

This script will export allele sequences in Extended Multi-FASTA (XMFA) format suitable for loading into third-party applications, such as ClonalFrame. It will also produce concatenated FASTA files. Only DNA loci that have a corresponding database containing allele sequence identifiers, or DNA and peptide loci with genome sequences tagged, can be included. Please check the loci that you would like to include. Alternatively select one or more schemes to include all loci that are members of the scheme. If a sequence does not exist in the remote database, it will be replaced with gap characters. Aligned output is limited to 200 records. Please be aware that if you select the alignment option it may take a long time to generate the output file.

Select ids: Paste in list of ids to include, start a new line for each. Leave blank to include all ids.

Include in identifier: isolate, country, region, year, epidemiological year, age yr, age mth, sex, disease, source

Loci: 16S_rDNA, 16S_rRNA (SSU_rRNA), 23S_rRNA, abcZ, abcZ (NEIS1015), acnF (NEIS1279), ackA2 (NEIS1727), acnA (NEIS1729)

Schemes: Genetic information processing, Metabolism, Pilin, Typing, MLST, Finotyping antigens, 16S, Antigen genes

Options: If both allele designations and tagged sequences exist for a locus, choose how you want these handled: ☐ Submit

☒ Use sequences tagged from the bin
☐ Use allele sequence retrieved from external database

☒ Do not include sequences with problem flagged (defined alleles will still be used)
☒ Do not include incomplete sequences
 Include: 0 bp flanking sequence ☐

☐ Align sequences
 Aligner: MAFFT
☐ Translate sequences
☐ Concatenate in frame

Click submit.

PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Export allele sequences in XMFA/concatenated FASTA formats - Neisseria PubMLST

This analysis has been submitted to the job queue.

Please be aware that this job may take a long time depending on the number of sequences to align and how busy the server is. Alignment of hundreds of sequences can take many hours!

[Follow the progress of this job and view the output.](#)

Please note that the % complete value will only update after the alignment of each locus.

The job will be submitted to the job queue. Click the link to follow the progress and download the resulting files.

Sequences will be export in XMFA and FASTA file formats.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Job status viewer

Status

| | |
|--------------|------------------------------|
| Job id: | BIGSdb_29565_1405601815_9307 |
| Submit time: | 2014-07-17 13:56:55 |
| Status: | finished |
| Start time: | 2014-07-17 13:57:10 |
| Progress: | 100% |
| Stop time: | 2014-07-17 13:57:23 |
| Total time: | 12 seconds |

Output

- XMFA output file (not aligned)
- Concatenated FASTA (not aligned)
- Tar file containing output files

Please note that job results will remain on the server for 7 days.

13.2.1 Aligning sequences

By default, sequences will be exported unaligned - this is very quick since no processing is required. You can choose to align the sequences by checking the 'Align sequences' checkbox.

Options

If both allele designations and tagged sequences exist for a locus, choose how you want these handled: i

☒ Use sequences tagged from the bin
☐ Use allele sequence retrieved from external database

☒ Do not include sequences with problem flagged (defined alleles will still be used)
☒ Do not include incomplete sequences

Include 0 bp flanking sequence i

☒ Align sequences
 Aligner: MAFFT

☐ Translate sequences
☐ Concatenate in frame

You can also choose to use MUSCLE or MAFFT as the aligner. MAFFT is the default choice and is usually much quicker than MUSCLE. Both produce comparable results.

13.3 Contig export

The contig export plugin can be accessed by clicking the 'Contigs' link in the Export section of the contents page of isolate databases.

Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For every allelic profile in the profiles/sequence definition database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.

Query database

- Search database - advanced queries.
- Browse database - peruse all records.
- Search by combinations of loci (profiles) - including partial matching.
- List query - find isolates by matching a field to an entered list.
- Projects - main projects defined in database.

Option settings

- Set general options - including isolate table field handling.
- Set display and query options for locus, schemes or scheme fields.

Submissions

- Manage submissions

General information

- Isolates: 34218
- Last updated: 2015-06-30
- Update history
- About BIGSdb

Breakdown

- Single field
- Two field
- Unique combinations
- Scheme and alleles
- Publications
- Sequence bin

Export

- Export dataset
- Contigs**
- Sequences - XMF / concatenated FASTA formats

Analysis

- Codon usage
- Presence/absence status of loci
- Genome comparator
- BLAST

Miscellaneous

- Description of database fields

Alternatively, it can be accessed following a query by clicking the 'Contigs' button in the Export section at the bottom of the results table.

| | | | | | | | | | | | | |
|----|-----------|-------------|----------------|------|------------------------------|------------------------|----|------|----------------------------|------|------|------|
| 5 | M00240227 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1100 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 6 | M00282207 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | W | 1101 | ST-22 complex | | | |
| 7 | 7891 | B54; Z1054 | Finland | 1975 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 20 | 9 | F3-1 |
| 8 | M00242007 | | UK | 2000 | invasive (unspecified/other) | Neisseria meningitidis | B | 1102 | ST-18 complex | | 14 | |
| 9 | 0021/84 | | Czech Republic | 1984 | invasive (unspecified/other) | Neisseria meningitidis | W | 114 | ST-22 complex | | | |
| 10 | 6748 | B73; Z1073 | Canada | 1971 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 18-1 | 3 | F5-1 |
| 11 | 129 | B92; Z1092 | Germany | 1964 | invasive (unspecified/other) | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F3-6 |
| 12 | 0090/89 | | Czech Republic | 1989 | invasive (unspecified/other) | Neisseria meningitidis | B | 1015 | ST-32 complex/ET-5 complex | 7 | 16 | |
| 13 | 139M | B99; Z1099 | Philippines | 1968 | | Neisseria meningitidis | A | 1 | ST-1 complex/subgroup I/II | 5-2 | 10 | F5-1 |
| 14 | 0120/95 | | Czech Republic | 1995 | invasive (unspecified/other) | Neisseria meningitidis | X | 117 | | | 14 | |
| 15 | 1 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 864 | | | | |
| 16 | 2 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 854 | ST-18 complex | | | |
| 17 | 3 | | Germany | 1999 | carrier | Neisseria meningitidis | W | 174 | ST-174 complex | | | |
| 18 | 4 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 19 | ST-18 complex | | | |
| 19 | S3131 | B213; Z1213 | Ghana | 1973 | invasive (unspecified/other) | Neisseria meningitidis | A | 4 | ST-4 complex/subgroup IV | 7 | 13-1 | F1-5 |
| 20 | 5 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 21 | 6 | | Germany | 1999 | carrier | Neisseria meningitidis | NG | 198 | ST-198 complex | | | |
| 22 | 7 | | Germany | 1999 | carrier | Neisseria meningitidis | E | 60 | ST-60 complex | | | |
| 23 | 8 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 32 | ST-32 complex/ET-5 complex | | | |
| 24 | S4355 | B227; Z1227 | Denmark | 1974 | invasive (unspecified/other) | Neisseria meningitidis | A | 5 | ST-5 complex/subgroup III | 5-1 | 9 | F3-1 |
| 25 | 9 | | Germany | 1999 | carrier | Neisseria meningitidis | B | 930 | ST-334 complex | | | |

Analysis tools:

Breakdown:

Analysis:

Export:

Page:

Select the isolates for which you wish to export contig data for. If the export function was accessed following a query, isolates returned in the query will be pre-selected.

PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions

Help Toggle:

Contig analysis and export

Please select the required isolate ids from which contigs are associated - use Ctrl or Shift to make multiple selections. Please note that the total length of tagged sequence is calculated by adding up the length of all loci tagged within the contig - if these loci overlap then the total tagged length will be reported as being longer than it really is but it won't exceed the length of the contig.

Isolates

- 1) A4/M1027
- 2) 120M
- 7) 7891
- 10) 6748
- 11) 129E
- 13) 139M
- 19) S3131
- 24) S4355

All None

Options

Identify contigs with \geq 0 % of sequence untagged

FASTA header line: original designation

Action

Reset Submit

Filter by

Sequence method:

Project:

Experiment:

Minimum length:

At its simplest, press submit.

A table will be produced with download links. Clicking these will produce the contigs in FASTA format.

Please select the required isolate ids from which contigs are associated - use Ctrl or Shift to make multiple selections. Please note that the total length of tagged sequence is calculated by adding up the length of all loci tagged within the contig - if these loci overlap then the total tagged length will be reported as being longer than it really is but it won't exceed the length of the contig.

Isolates

- 1) A4/M1027
- 2) 120M
- 7) 7891
- 10) 6748
- 11) 129E
- 13) 139M
- 19) S3131
- 24) S4355

All None

Options

Identify contigs with \geq 0 % of sequence untagged

FASTA header line: original designation

Action

Reset Submit

Filter by

Sequence method:

Project:

Experiment:

Minimum length:

Contigs with \geq 0% sequence length untagged

| id | isolate | contigs | matching contigs | | non-matching contigs | |
|----|----------|---------|------------------|----------|----------------------|----------|
| | | | count | download | count | download |
| 1 | A4/M1027 | 364 | 364 | | 0 | |
| 2 | 120M | 359 | 359 | | 0 | |
| 7 | 7891 | 199 | 199 | | 0 | |
| 10 | 6748 | 652 | 652 | | 0 | |

- Download table in tab-delimited text format
- Batch download all contigs from selected isolates (tar format)

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

Please select the required isolate ids from which contigs are associated - use Ctrl or Shift to make multiple selections. Please note that the total length of tagged sequence is calculated by adding up the length of all loci tagged within the contig - if these loci overlap then the total tagged length will be reported as being longer than it really is but it won't exceed the length of the contig.

Isolates
1) A4/M1027
2) 120M
7) 7891
10) 6748
11) 129E
13) 139M
19) S3131
24) S4355
All None

Options
Identify contigs with \geq 0 % of sequence untagged
FASTA header line: original designation
Reset Submit

Filter by
Sequence method:
Project:
Experiment:
Minimum length:

Contigs with $\geq 0\%$ sequence length untagged

| id | isolate | contigs | matching contigs | | non-matching contigs | |
|----|----------|---------|------------------|----------|----------------------|----------|
| | | | count | download | count | download |
| 1 | A4/M1027 | 364 | 364 | | 0 | |
| 2 | 120M | 359 | 359 | | 0 | |
| 7 | 7891 | 199 | 199 | | 0 | |
| 10 | 6748 | 652 | 652 | | 0 | |

- Download table in tab-delimited text format
- Batch download all contigs from selected isolates (tar format)

13.3.1 Filtering by tagged status of contigs

You can also export contigs based on the percentage of the sequence that has been tagged. This is useful to find sequences to target for gene discovery.

In order to export contigs where at least half the sequence has been tagged (and also the remaining contigs in a separate file), select '50' in the dropdown box for %untagged.

Options

Identify contigs with \geq 50 % of sequence untagged

FASTA header line: original designation

The resulting table has two download links for each isolate, one for contigs matching the condition, and one for contigs that don't match.

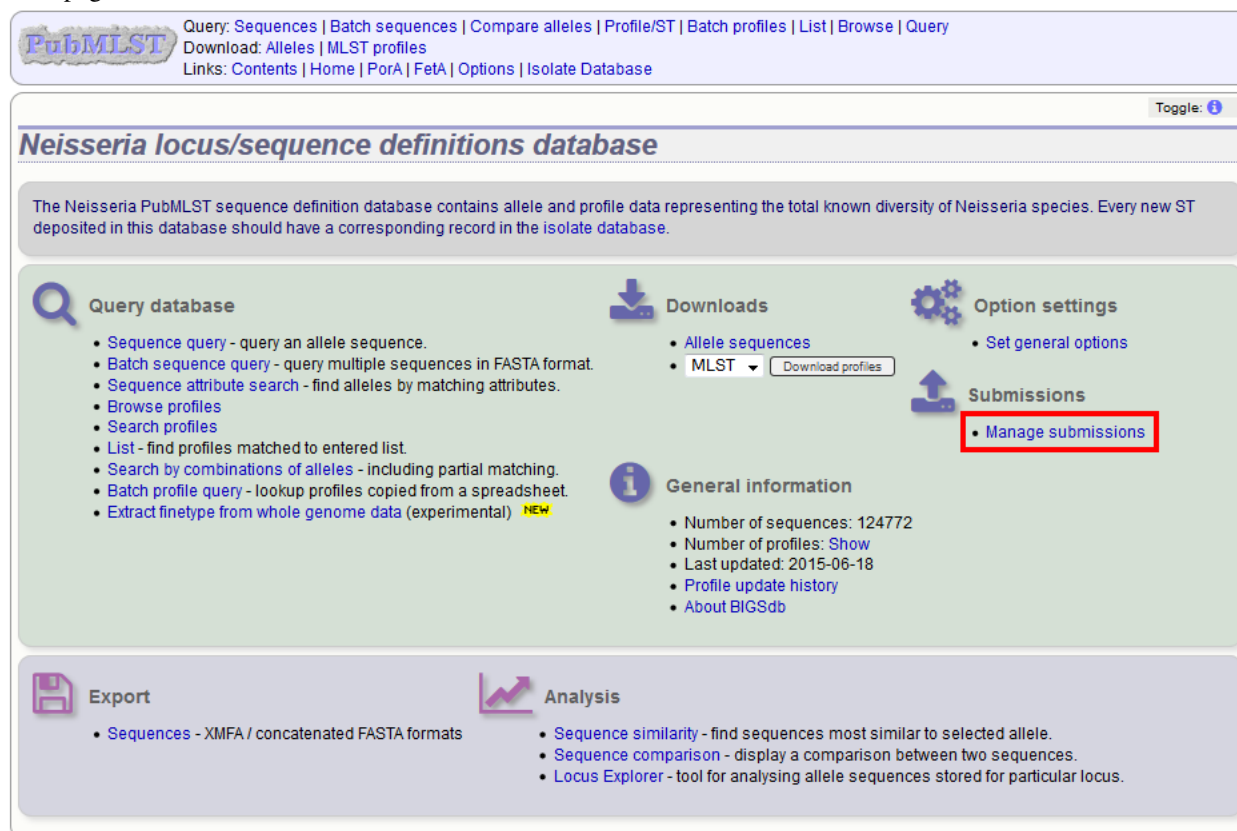
Contigs with $\geq 50\%$ sequence length untagged

| id | isolate | contigs | matching contigs | | non-matching contigs | |
|----|----------|---------|------------------|----------|----------------------|----------|
| | | | count | download | count | download |
| 1 | A4/M1027 | 364 | 163 | | 201 | |
| 2 | 120M | 359 | 81 | | 278 | |
| 7 | 7891 | 199 | 48 | | 151 | |
| 10 | 6748 | 652 | 393 | | 259 | |

Submitting data using the submission system

The automated submission system allows users to submit data (new alleles, profiles, or isolates) to the database curators for assignment and upload to the database. The submission system is enabled on a per-database basis so will not always be available.

If the system is enabled, new submissions can be made by clicking the ‘Manage submissions’ link on the database front page.



The screenshot shows the Neisseria locus/sequence definitions database interface. At the top, there is a navigation bar with links for Query, Sequences, Batch sequences, Compare alleles, Profile/ST, Batch profiles, List, Browse, and Query. Below this, there are links for Download: Alleles, MLST profiles, and Links: Contents, Home, PorA, FetA, Options, and Isolate Database. The main heading is 'Neisseria locus/sequence definitions database'. A description states: 'The Neisseria PubMLST sequence definition database contains allele and profile data representing the total known diversity of Neisseria species. Every new ST deposited in this database should have a corresponding record in the isolate database.' The interface is divided into several sections: 'Query database' with links for Sequence query, Batch sequence query, Sequence attribute search, Browse profiles, Search profiles, List, Search by combinations of alleles, Batch profile query, and Extract finetype from whole genome data; 'Downloads' with links for Allele sequences and MLST; 'Option settings' with a link for Set general options; 'Submissions' with a link for Manage submissions (highlighted with a red box); 'General information' with statistics on sequences, profiles, and last updated date; 'Export' with a link for Sequences; and 'Analysis' with links for Sequence similarity, Sequence comparison, and Locus Explorer.

14.1 Registering a user account

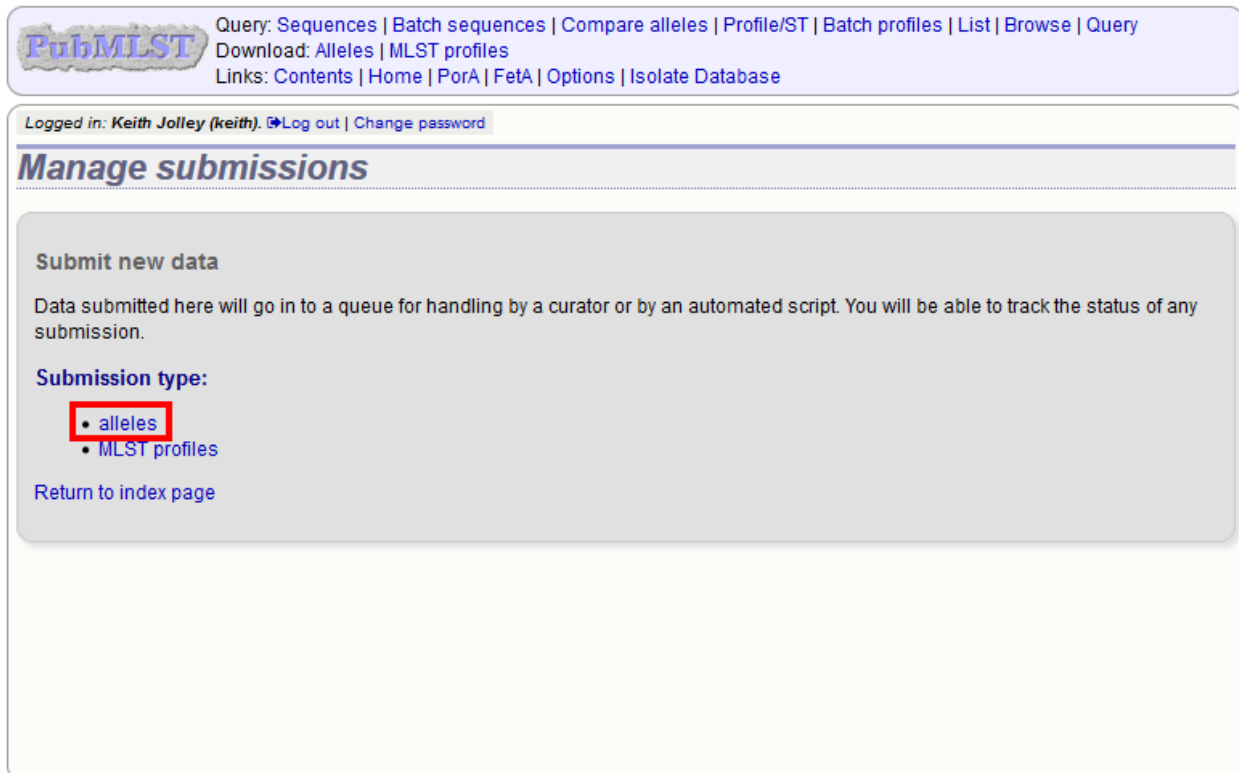
You must have an account for the appropriate database in order to use the submission system. This will need to be set up by a curator, so contact them in the first instance.

14.2 Allele submission

New allele data can only be submitted from within the appropriate sequence definition database. Submissions consist of one or more new allele sequences for a single locus. You will need to create separate submissions for each locus - this is because different loci may be handled by different curators.

14.2.1 Start

Click the ‘alleles’ link under submission type on the submission management page.



PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#)

Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- **alleles**
- MLST profiles

[Return to index page](#)

14.2.2 Select the submission locus

Select the locus from the locus list box:

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

Manage submissions

Submit new alleles

You need to make a separate submission for each locus for which you have new alleles - this is because different loci may have different curators. You can submit any number of new sequences for a single locus as one submission. Sequences should be trimmed to the correct start/end sites for the selected locus.

Filter loci by scheme

- All loci
- Capsule
- Genetic Information Processing
- Metabolism
- Typing
- Other schemes
- Loci not in schemes

Select locus

- abcZ
- abcZ (NEIS1015)
- aceF (NEIS1279)
- acnA (NEIS1729)
- acnB (NEIS1492)
- adk
- adk (NEIS0767)
- aroE
- aroE (NEIS1810)

Sequence details

technology: Illumina
read length: 100-199
coverage: 20-49x
assembly: de novo
assembly software: Velvet

FASTA or single sequence

Action
Submit

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'.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

Manage submissions

Submit new alleles

You need to make a separate submission for each locus for which you have new alleles - this is because different loci may have different curators. You can submit any number of new sequences for a single locus as one submission. Sequences should be trimmed to the correct start/end sites for the selected locus.

Filter loci by scheme

- All loci
- Capsule
- Genetic Information Processing
- Metabolism
- Typing
- Other schemes
- Loci not in schemes

Select locus

- abcZ
- abcZ (NEIS1015)
- aceF (NEIS1279)
- acnA (NEIS1729)
- acnB (NEIS1492)
- adk
- adk (NEIS0767)
- aroE
- aroE (NEIS1810)

Sequence details

technology: Illumina
read length: 100-199
coverage: 20-49x
assembly: de novo
assembly software: Velvet

FASTA or single sequence

Action
Submit

The locus list is now constrained making selection easier.

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
 Download: [Alleles](#) | [MLST profiles](#)
 Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submit new alleles

You need to make a separate submission for each locus for which you have new alleles - this is because different loci may have different curators. You can submit any number of new sequences for a single locus as one submission. Sequences should be trimmed to the correct start/end sites for the selected locus.

Filter loci by scheme

- All loci
- Capsule
- Genetic Information Processing
- Metabolism
- Typing
 - MLST
 - Finotyping antigens
 - Antigen genes
 - eMLST (20 locus partial genes)
 - eMLST (20 locus whole genes)
 - Other schemes
 - Loci not in schemes

Filter

Select locus

- abcZ
- adk
- aroE
- fumC
- gdh
- pdhC
- pgm

Sequence details

technology:

read length:

coverage:

assembly:

assembly software:

FASTA or single sequence

Action

Submit

14.2.3 Enter details of sequencing method

There are a number of fields that must be filled in so that the curator knows how the sequence was obtained:

- technology - the 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.

14.2.4 Paste in sequence(s)

Paste in the new variant sequences to the box. This can either be a stand- alone sequence or multiple sequences in FASTA format. The sequences must be trimmed to the start and end points of the loci - check existing allele definitions if in doubt. The submission is likely to be rejected if sequences are not trimmed. Click submit.

Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
 Download: [Alleles](#) | [MLST profiles](#)
 Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: Keith Jolley (keith) | [Log out](#) | [Change password](#)

Manage submissions

Submit new alleles

You need to make a separate submission for each locus for which you have new alleles - this is because different loci may have different curators. You can submit any number of new sequences for a single locus as one submission. Sequences should be trimmed to the correct start/end sites for the selected locus.

Filter loci by scheme

- All loci
- Capsule
- Genetic Information Processing
- Metabolism
- Typing
 - ☒ MLST
 - Finetyping antigens
 - Antigen genes
 - eMLST (20 locus partial genes)
 - eMLST (20 locus whole genes)
 - Other schemes
 - Loci not in schemes

Select locus

Sequence details


technology:
 read length:
 coverage:
 assembly:
 assembly software:

FASTA or single sequence

```
>NM322
TTTGATACTGTCGCCGAAGGTTTGGGCGAAATTCGCGATTAT
TGCGCCGTTATCATCATGTCAGCCATGAGTTGAAATGGTTC
GAGTGAGGCCCTTATTGAAGAGCTCAACGAATTGCAACTTGAG
ATCGAAGCGAAGGACCGGCTGGAAGTTGGATCGGCGGTGAAC
AGACTTTGGGCGAACTCGGTTTGGCGAAACGAAAAATCGG
CAACCTCTCCGCGCTCAGAAAAAGCGGTCGCTTGGGCGAG
```

Action

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.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Error:
 Sequence 'NM432' has already been defined as abcZ-3.

Submit new alleles
 You need to make a separate submission for each locus for which you have new alleles - this is because different loci may have different curators. You can submit any number of new sequences for a single locus as one submission. Sequences should be trimmed to the correct start/end sites for the selected locus.

Filter loci by scheme

- All loci
- Capsule
- Genetic Information Processing
- Metabolism
- Typing
 - MLST
 - Finotyping antigens
 - Antigen genes
 - eMLST (20 locus partial genes)
 - eMLST (20 locus whole genes)
- Other schemes
- Loci not in schemes

Select locus
 Filter


- abcZ
- adk
- aroE
- fumC
- gdh
- pdhC
- pgm

Sequence details
 technology:
 read length:
 coverage:
 assembly:
 assembly software:

FASTA or single sequence
 >NM322
 TTGTGATCTGTCGCCGAAGGTTTGGGCGAA
 ATTGCGGATTTATTGCGCGTTATCATCAT
 GTCAGCCATGAGTTGGAAATGGTTTCGAGT
 GAGGCCTTATTGAAAGAGCTCAACGAATTG
 CAACTTGAGATCGAAGCGAAGGACGGCTGG

Action


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


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150619110123_31937_75843

Sequences
 You are submitting the following abcZ sequences: [Download](#) 

| Identifier | Length | Sequence | Status | Assigned allele |
|------------|--------|---|---------|-----------------|
| NM322 | 433 | TTGTGATCTGTCGCCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |
| NM21 | 433 | TTGTGATCCGTTGCCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |

Sequence details
 technology:
 read length:
 coverage:
 assembly:
 assembly software:

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

Action

Supporting files
 Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

No files selected.

Messages

14.2.5 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click ‘Add message’.


PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
 Download: [Alleles](#) | [MLST profiles](#)
 Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150619110123_31937_75843

Sequences

You are submitting the following abcZ sequences: [Download](#) 

| Identifier | Length | Sequence | Status | Assigned allele |
|------------|--------|---|---------|-----------------|
| NM322 | 433 | TTTGATACGTGCGCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |
| NM21 | 433 | TTTGATACCGTTGCGGAAGG ... GCGGATTGTCGAACTCGATC | pending | |

Sequence details

technology: Illumina
 read length: 100-199
 coverage: 20-49x
 assembly: de novo
 assembly software: Velvet

E-mail

Updates will be sent to keith.jolley@zoo.ox.ac.uk.

☒ E-mail submission updates

Action

[Finalize submission!](#)

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected. [Upload files](#)

Messages

The sequence variant NM322 has been seen in 3 isolates.

[Add message](#)


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.

PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150619110123_31937_75843

Sequences
You are submitting the following abcZ sequences: [Download](#) 

| Identifier | Length | Sequence | Status | Assigned allele |
|------------|--------|---|---------|-----------------|
| NM322 | 433 | TTTGATACTGTCGCCGAAGG ... GCGGATTGTGCGAACTCGATC | pending | |
| NM21 | 433 | TTTGATACCGTTCGCCGAAGG ... GCGGATTGTGCGAACTCGATC | pending | |

Sequence details
technology: Illumina
read length: 100-199
coverage: 20-49x
assembly: de novo
assembly software: Velvet

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

Action
[Finalize submission!](#)

Supporting files
Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.
[Browse...](#) No files selected. [Upload files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-19 10:10:04+00 | Keith Jolley | The sequence variant NM322 has been seen in 3 isolates. |

[Add message](#)

14.2.6 Add supporting files

Some submissions will require the attachment of supporting files. This will depend on the policies of the individual databases. Sequences determined by Sanger sequencing should normally have forward and reverse trace files attached.

Files can be added to the submission by clicking the 'Browse' button in the 'Supporting files' section.


Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150619110123_31937_75843

Sequences

You are submitting the following abcZ sequences: [Download](#) 

| Identifier | Length | Sequence | Status | Assigned allele |
|------------|--------|---|---------|-----------------|
| NM322 | 433 | TTTGATACTGTCGCCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |
| NM21 | 433 | TTTGATACCGTTGCCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |

Sequence details

technology: Illumina
 read length: 100-199
 coverage: 20-49x
 assembly: de novo
 assembly software: Velvet

E-mail

Updates will be sent to keith.jolley@zoo.ox.ac.uk.

☒ E-mail submission updates

Action

[Finalize submission!](#)

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected. [Upload files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-19 10:10:04+00 | Keith Jolley | The sequence variant NM322 has been seen in 3 isolates. |

[Add message](#)

Select the file in the selection box, then click 'Upload files'.


Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150619110123_31937_75843

Sequences

You are submitting the following abcZ sequences: [Download](#) 

| Identifier | Length | Sequence | Status | Assigned allele |
|------------|--------|---|---------|-----------------|
| NM322 | 433 | TTTGATACTGTCGCCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |
| NM21 | 433 | TTTGATACCGTTGCCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |

Sequence details

technology: Illumina
 read length: 100-199
 coverage: 20-49x
 assembly: de novo
 assembly software: Velvet

E-mail

Updates will be sent to keith.jolley@zoo.ox.ac.uk.

☒ E-mail submission updates

Action

[Finalize submission!](#)

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) abcZ.ace [Upload files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-19 10:10:04+00 | Keith Jolley | The sequence variant NM322 has been seen in 3 isolates. |


[Add message](#)

The file will be uploaded and shown in a table.

manage submissions

Submission: BIGSdb_20150619110123_31937_75843

Sequences

You are submitting the following abcZ sequences: [Download](#) 

| Identifier | Length | Sequence | Status | Assigned allele |
|------------|--------|---|---------|-----------------|
| NM322 | 433 | TTTGATACTGTCGCCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |
| NM21 | 433 | TTTGATACCGTTGCCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |

Sequence details

technology: Illumina
 read length: 100-199
 coverage: 20-49x
 assembly: de novo
 assembly software: Velvet

E-mail

Updates will be sent to keith.jolley@zoo.ox.ac.uk.

☒ E-mail submission updates

Action

[Finalize submission!](#)

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected. [Upload files](#)

| Filename | Size | Delete |
|----------|--------|--------------------------|
| abcZ.ace | 1.7 MB | <input type="checkbox"/> |

[Delete selected files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-19 10:10:04+00 | Keith Jolley | The sequence variant NM322 has been seen in 3 isolates. |

[Add message](#)

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

14.2.7 Finalize submission

Make sure the ‘E-mail submission updates’ box is checked if you wish to receive E-mail notification of the result of your submission. This setting is remembered between submissions.

Click ‘Finalize submission!’.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | Feta | Options | Isolate Database

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

Manage submissions

Submission: BIGSdb_20150619110123_31937_75843

Sequences

You are submitting the following abcZ sequences: Download

| Identifier | Length | Sequence | Status | Assigned allele |
|------------|--------|--|---------|-----------------|
| NM322 | 433 | TTTGATACTGTGCGCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |
| NM21 | 433 | TTTGATACCGTTGCGCGAAGG ... GCGGATTGTCGAACTCGATC | pending | |

Sequence details

technology: Illumina
read length: 100-199
coverage: 20-49x
assembly: de novo
assembly software: Velvet

E-mail

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

Action
Finalize submission!

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

Browse... No files selected. Upload files

| Filename | Size | Delete |
|----------|--------|--------------------------|
| abcZ ace | 1.7 MB | <input type="checkbox"/> |

Delete selected files

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-19 10:10:04+00 | Keith Jolley | The sequence variant NM322 has been seen in 3 isolates. |

Add message

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

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | List | Browse | Query
Download: Alleles | MLST profiles
Links: Contents | Home | PorA | Feta | Options | Isolate Database

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

Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- alleles
- MLST profiles

Pending submissions

You have submitted the following submissions that are pending curation:

| Submission id | Submitted | Updated | Type | Details |
|-----------------------------------|------------|------------|---------|------------------|
| BIGSdb_20150619110123_31937_75843 | 2015-06-19 | 2015-06-19 | alleles | 2 abcZ sequences |

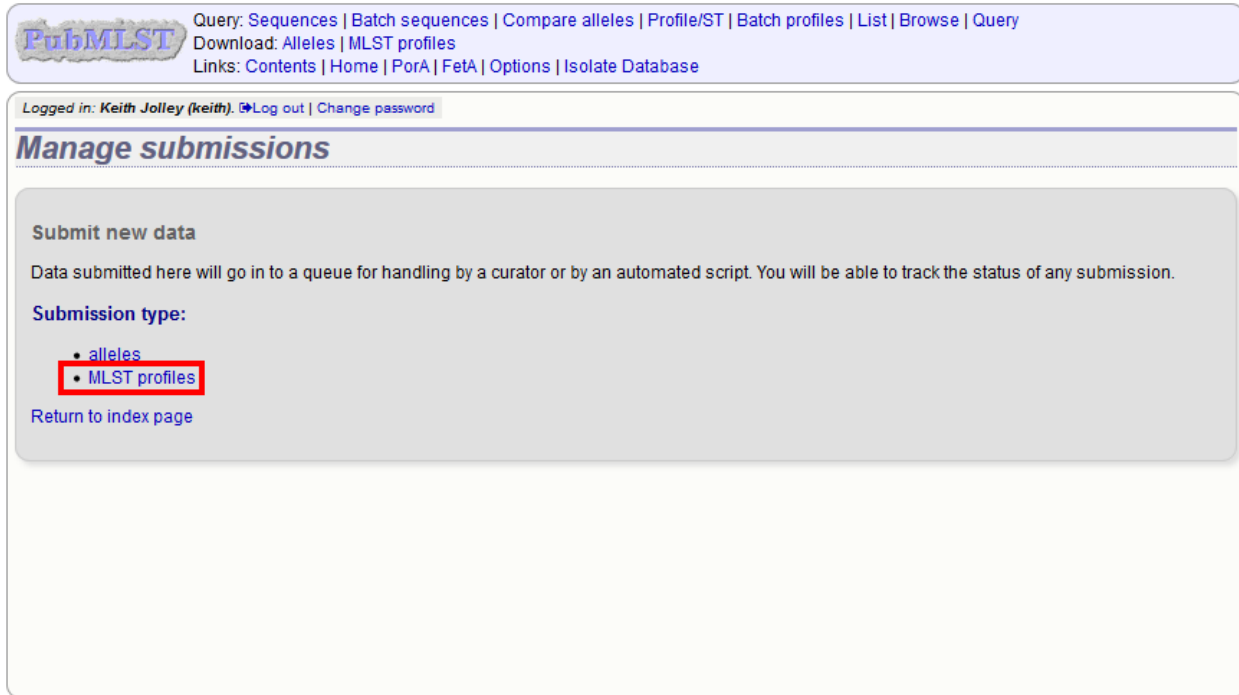
Return to index page

14.3 Profile submission

Even if the submission system has been enabled, submission of profiles has to be specifically enabled. Many databases require submission of representative isolate data for any new allelic profile for schemes such as MLST. In this case any new profiles can be readily extracted from the isolate data so profiles do not need to be submitted separately.

14.3.1 Start

Click the appropriate profiles link under submission type on the submission management page.



The screenshot shows the PubMLST submission management interface. At the top, there is a navigation bar with the PubMLST logo and links for Query (Sequences, Batch sequences, Compare alleles, Profile/ST, Batch profiles, List, Browse, Query), Download (Alleles, MLST profiles), and Links (Contents, Home, PorA, Feta, Options, Isolate Database). Below this, a status bar indicates the user is logged in as Keith Jolley (keith) with links to Log out and Change password. The main heading is "Manage submissions". Under the "Submit new data" section, a message states that data submitted here will go into a queue for handling by a curator or by an automated script. The "Submission type:" section lists two options: "alleles" and "MLST profiles", with "MLST profiles" highlighted by a red box. A "Return to index page" link is located at the bottom of the submission type section.

Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [Feta](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submit new data


Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- [alleles](#)
- [MLST profiles](#)

[Return to index page](#)

Download the Excel submission template.

Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submit new MLST profiles

Paste in your profiles for assignment using the template available below.

- [Download tab-delimited header for your spreadsheet](#) - use 'Paste Special' > 'Text' to paste the data.
- [Download submission template \(xlsx format\)](#)


Please paste in tab-delimited text (include a field header line)

Action
Submit

14.3.2 Paste in profile(s)

Fill in the template. The first column 'id' can be used to enter an identifier that is meaningful to you - it is used to report back the results but is not uploaded to the database. It can be left blank, or the entire column can be removed - in which case individual profiles will be identified by row number.

Copy and paste the entire contents of the submission worksheet. Click submit.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submit new MLST profiles

Paste in your profiles for assignment using the template available below.


- [Download tab-delimited header for your spreadsheet](#) - use 'Paste Special' > 'Text' to paste the data.
- [Download submission template \(xlsx format\)](#)

Please paste in tab-delimited text (include a field header line)

| id | abcZ | adk | aroE | fumC | gdh | pdhC | pgm |
|----|------|-----|------|------|-----|------|-----|
| 8 | 5 | 32 | 2 | 6 | 43 | 32 | |
| 5 | 7 | 3 | 6 | 33 | 12 | 2 | |
| 3 | 2 | 4 | 3 | 8 | 4 | 6 | |

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.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Error:
 Row 3: Profile has already been defined as ST-10018.

Submit new MLST profiles
 Paste in your profiles for assignment using the template available below.

- Download tab-delimited header for your spreadsheet - use 'Paste Special' ➡ Text to paste the data.
- Download submission template (xlsx format)

Please paste in tab-delimited text (include a field header line)


| id | abcZ | adk | aroE | fumC | gdh | pdhC | pgm |
|----|------|-----|------|------|-----|------|-----|
| 8 | 5 | 32 | 2 | 6 | 43 | 32 | |
| 5 | 7 | 3 | 6 | 33 | 12 | 2 | |
| 3 | 2 | 4 | 3 | 8 | 4 | 6 | |

Action

Provided the checks pass, you will then be able to add additional information to your submission

14.3.3 Add message to curator


If you wish to enter a message to the curator, enter this in the messages box and click 'Add message'.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150619135509_21808_97411

Profiles
 You are submitting the following MLST profiles: [Download](#) 

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

Action
[Finalize submission!](#)

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|---------|-------------|
| Row 1 | 5 | 8 | 32 | 2 | 6 | 43 | 32 | pending | |
| Row 2 | 7 | 5 | 3 | 6 | 33 | 12 | 2 | pending | |

Supporting files
 Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.


[Browse...](#) No files selected. [Upload files](#)

Messages

Corresponding isolate data has already been submitted (ids 43721-43722).

[Add message](#)


The message will be attached. A curator may respond to the message and attach their own, with the full conversation becoming part of the submission record.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150619135509_21808_97411

Profiles
 You are submitting the following MLST profiles: [Download](#) 

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

Action
[Finalize submission!](#)

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|---------|-------------|
| Row 1 | 5 | 8 | 32 | 2 | 6 | 43 | 32 | pending | |
| Row 2 | 7 | 5 | 3 | 6 | 33 | 12 | 2 | pending | |

Supporting files
 Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected. [Upload files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|--|
| 2015-06-19 13:00:33+00 | Keith Jolley | Corresponding isolate data has already been submitted (ids 43721-43722). |

[Add message](#)

14.3.4 Add supporting files

Some submissions may require the attachment of supporting files. These files can be added to the submission by clicking the 'Browse' button in the 'Supporting files' section.

Select the file in the selection box, then click 'Upload files'.

14.3.5 Finalize submission

Make sure the 'E-mail submission updates' box is checked if you wish to receive E-mail notification of the result of your submission. This setting is remembered between sessions.

Click 'Finalize submission!'.

Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
 Download: [Alleles](#) | [MLST profiles](#)
 Links: [Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submission: **BIGSdb_20150619135509_21808_97411**

Profiles
 You are submitting the following MLST profiles: [Download](#)

| Identifier | adk | abcZ | aroE | fumC | gdh | pdhC | pgm | Status | Assigned ST |
|------------|-----|------|------|------|-----|------|-----|---------|-------------|
| Row 1 | 5 | 8 | 32 | 2 | 6 | 43 | 32 | pending | |
| Row 2 | 7 | 5 | 3 | 6 | 33 | 12 | 2 | pending | |

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

Action
[Finalize submission!](#)


Supporting files
 Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.
[Browse...](#) No files selected. [Upload files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|--|
| 2015-06-19 13:00:33+00 | Keith Jolley | Corresponding isolate data has already been submitted (ids 43721-43722). |

[Add message](#)

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


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [PorA](#) | [FetA](#) | [Options](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- [alleles](#)
- [MLST profiles](#)

Pending submissions

You have submitted the following submissions that are pending curation:

| Submission id | Submitted | Updated | Type | Details |
|---|------------|------------|----------|-----------------|
| BIGSdb_20150619135509_21808_97411 | 2015-06-19 | 2015-06-19 | profiles | 2 MLST profiles |


[Return to index page](#)

14.4 Isolate submission

New isolate data can only be submitted from within the appropriate isolate database. You may be required to submit isolate data if you would like to get a new MLST sequence type defined, but this depends on individual database policy.

14.4.1 Start

Click the ‘isolates’ link under submission type on the submission management page.


Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submit new data


Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- isolates**

[Return to index page](#)

Download the Excel submission template.


Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submit new isolates

Paste in your isolates for addition to the database using the template available below.

- Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.
- Enter references for your isolates as a semi-colon (;) separated list of PubMed ids.
- You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name.
- [Download tab-delimited header for your spreadsheet](#) - use 'Paste Special' > 'Text' to paste the data.
- [Download submission template \(xlsx format\)](#)

Please paste in tab-delimited text (include a field header line)

Action

Submit

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.

Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

Logged in: Keith Jolley (keith) [Log out] [Change password]

Manage submissions

Submit new isolates

Paste in your isolates for addition to the database using the template available below.


- Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.
- Enter references for your isolates as a semi-colon (;) separated list of PubMed ids.
- You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name.
- [Download tab-delimited header for your spreadsheet](#) - use "Paste Special > Text" to paste the data.
- [Download submission template \(xlsx format\)](#)

Please paste in tab-delimited text (include a field header line)

| isolate | aliases | references | country | region | year | age_yr | age_mth | sex |
|------------------|-----------------------|------------------------|------------------------|------------------|------|-------------|---------|-------|
| disease | source | epidemiology | species | serogroup | MLEE | designation | | |
| serotype | sero_subtype | ET no | penicillin | penicillin_range | | | | |
| amoxicillin | sulphonamide | ceftriaxone | ceftriaxone_range | | | | | |
| chloramphenicol | chloramphenicol_range | cefotaxime | cefotaxime_range | | | | | |
| rifampicin | rifampicin_range | ciprofloxacin | ciprofloxacin_range | | | | | |
| pending_assembly | | abg2 | adk | aroE | fumC | gdh | | |
| pdhC | pgm | FetA VR | PorA VR1 | PorA VR2 | | | | |
| UK322 | | UK | 2015 | | | | | |
| meningitis and | septicaemia | blood | Neisseria meningitidis | | | | | |
| B | | | | | | | | |
| 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 2 |
| UK325 | | UK | 2015 | | | | | |
| septicaemia | CSF | Neisseria meningitidis | | | | | | |
| B | | | | | | | | |
| 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5-1 2 |

[Submit](#)

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.


Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Error:
UK325 has problems - species: 'Neisseria meningitis' 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.

- Enter aliases (alternative names) for your isolates as a semi-colon (;) separated list.
- Enter references for your isolates as a semi-colon (;) separated list of PubMed ids.
- You can also upload additional allele fields along with the other isolate data - simply create a new column with the locus name.
- [Download tab-delimited header for your spreadsheet](#) - use 'Paste Special' > 'Text' to paste the data.
- [Download submission template \(xlsx format\)](#)

Please paste in tab-delimited text (include a field header line)

isolate aliases references country region year age_yr age_mth sex
disease source epidemiology species serogroup MLEE_designation
serotype sero_subtype ET_no penicillin penicillin_range
amoxicillin sulphonamide ceftriaxone ceftriaxone_range
chloramphenicol chloramphenicol_range cefotaxime cefotaxime_range
rifampicin rifampicin_range ciprofloxacin ciprofloxacin_range
pending_assembly comments abcZ adk aroE fumC gdh
pdhC pgm FetA_VR PorA_VR1 PorA_VR2
UK322 UK 2015
meningitis and septicaemia blood Neisseria meningitidis
B
2 3 4 3 8 4 6 F1-5 5 2
UK325 UK 2015
septicaemia CSF Neisseria meningitis
B
2 3 4 3 8 4 6 F1-5 5-1 2

Action
Submit

Provided the checks pass, you will then be able to add additional information to your submission.

14.4.3 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Add message'.


PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

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

Manage submissions

Submission: BIGSdb_20150622133440_1412_50186

Isolates

You are submitting the following isolates: [Download](#) 

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK322 | UK | 2015 | meningitis and septicaemia | blood | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK325 | UK | 2015 | septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 18 | 4 | 6 | F1-5 | 5-1 | 2 |

E-mail

Updates will be sent to keith.jolley@zoo.ox.ac.uk.

☐ E-mail submission updates

Action

[Finalize submission!](#)

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected. [Upload files](#)

Messages

I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.

[Add message](#)

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.


PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

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

Manage submissions

Submission: BIGSdb_20150622133440_1412_50186

Isolates

You are submitting the following isolates: [Download](#) 

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK322 | UK | 2015 | meningitis and septicaemia | blood | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK325 | UK | 2015 | septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 18 | 4 | 6 | F1-5 | 5-1 | 2 |

E-mail

Updates will be sent to keith.jolley@zoo.ox.ac.uk.

☐ E-mail submission updates

Action

[Finalize submission!](#)

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected. [Upload files](#)

Messages


| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-22 12:36:41+00 | Keith Jolley | I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks. |

[Add message](#)

14.4.4 Add supporting files

You can add any files required to support the submission. You may, for example, wish to include a genome sequence for an isolate record (contigs in FASTA format). If you are doing this, make sure that the filename can be unambiguously linked to the appropriate isolate record and *add a message*.


Files can be added to the submission by clicking the ‘Browse’ button in the ‘Supporting files’ section.


[Query: Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
[Breakdown: Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
[Links: Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150622133440_1412_50186

Isolates
 You are submitting the following isolates: [Download](#) 

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK322 | UK | 2015 | meningitis and septicaemia | blood | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK325 | UK | 2015 | septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 18 | 4 | 6 | F1-5 | 5-1 | 2 |

E-mail
 Updates will be sent to keith.jolley@zoo.ox.ac.uk. [Finalize submission!](#)
☐ E-mail submission updates

Supporting files
 Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected.
 [Upload files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-22 12:36:41+00 | Keith Jolley | I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks. |
| | | I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks. |

[Add message](#)

Select the file in the selection box, then click ‘Upload files’.


PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

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

Manage submissions

Submission: BIGSdb_20150622133440_1412_50186

Isolates

You are submitting the following isolates: [Download](#) 

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK322 | UK | 2015 | meningitis and septicaemia | blood | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK325 | UK | 2015 | septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 18 | 4 | 6 | F1-5 | 5-1 | 2 |

E-mail

Updates will be sent to keith.jolley@zoo.ox.ac.uk.

☐ E-mail submission updates

Action

[Finalize submission!](#)

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) UK325_genome_contigs.fas [Upload files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-22 12:36:41+00 | Keith Jolley | I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks. |
| | | I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks. |

[Add message](#)

The file will be uploaded and shown in a table.


PubMLST Query: Search | Browse | Profile/ST | List
Breakdown: Isolate fields | Scheme/alleles | Publications
Links: Contents | Home | Options | Profiles/sequences definitions | Database submissions

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

Manage submissions

Submission: BIGSdb_20150622133440_1412_50186

Isolates

You are submitting the following isolates: [Download](#) 

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK322 | UK | 2015 | meningitis and septicaemia | blood | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK325 | UK | 2015 | septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 18 | 4 | 6 | F1-5 | 5-1 | 2 |

E-mail

Updates will be sent to keith.jolley@zoo.ox.ac.uk.

☐ E-mail submission updates

Action

[Finalize submission!](#)

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected. [Upload files](#)

| Filename | Size | Delete |
|--------------------------|------|--------------------------|
| UK325_genome_contigs.fas | 2 MB | <input type="checkbox"/> |

[Delete selected files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-22 12:36:41+00 | Keith Jolley | I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks. |

[Add message](#)

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

14.4.5 Finalize submission

Make sure the ‘E-mail submission updates’ box is checked if you wish to receive E-mail notification of the result of your submission. This setting is remembered between sessions.

Click ‘Finalize submission!’.


PubMLST Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
 Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
 Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)

Manage submissions

Submission: BIGSdb_20150622133440_1412_50186

Isolates

You are submitting the following isolates: [Download](#) 

| isolate | country | year | disease | source | species | serogroup | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | FetA_VR | PorA_VR1 | PorA_VR2 |
|---------|---------|------|----------------------------|--------|------------------------|-----------|------|-----|------|------|-----|------|-----|---------|----------|----------|
| UK322 | UK | 2015 | meningitis and septicaemia | blood | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 8 | 4 | 6 | F1-5 | 5 | 2 |
| UK325 | UK | 2015 | septicaemia | CSF | Neisseria meningitidis | B | 2 | 3 | 4 | 3 | 18 | 4 | 6 | F1-5 | 5-1 | 2 |

E-mail **Action**

Updates will be sent to keith.jolley@zoo.ox.ac.uk. Finalize submission!

☒ E-mail submission updates

Supporting files

Please upload any supporting files required for curation. Ensure that these are named unambiguously or add an explanatory note so that they can be linked to the appropriate submission item. Individual filesize is limited to 32 MB.

[Browse...](#) No files selected. [Upload files](#)

| Filename | Size | Delete |
|--------------------------|------|--------------------------|
| UK325_genome_contigs.fas | 2 MB | <input type="checkbox"/> |


[Delete selected files](#)

Messages

| Timestamp | User | Message |
|------------------------|--------------|---|
| 2015-06-22 12:36:41+00 | Keith Jolley | I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks. |

[Add message](#)

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

Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [Profiles/sequences definitions](#) | [Database submissions](#)

Logged in: [Keith Jolley \(keith\)](#) | [Log out](#) | [Change password](#)

Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- [isolates](#)

Pending submissions

You have submitted the following submissions that are pending curation:

| Submission id | Submitted | Updated | Type | Details |
|--|------------|------------|----------|------------|
| BIGSdb_20150622133440_1412_50186 | 2015-06-22 | 2015-06-22 | isolates | 2 isolates |

[Return to index page](#)

RESTful Application Programming Interface (API)

The REST API allows third-party applications to retrieve data stored within BIGSdb databases. 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 file).

15.1 Passing optional parameters

Optional parameters can be passed as arguments to the query URL by adding a ‘?’ followed by the first argument and its value (separated by a ‘=’). Additional parameters are separated by a ‘&’, e.g.

http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates?page=2&page_size=100

15.2 Resources

- */ or /db*
- */db/{database}*
- */db/{database}/loci*
- */db/{database}/loci/{locus}*
- */db/{database}/loci/{locus}/alleles*
- */db/{database}/loci/{locus}/alleles_fasta*
- */db/{database}/loci/{locus}/alleles/{allele_id}*
- */db/{database}/schemes*
- */db/{database}/schemes/{scheme_id}*
- */db/{database}/schemes/{scheme_id}/fields/{field}*
- */db/{database}/schemes/{scheme_id}/profiles*
- */db/{database}/schemes/{scheme_id}/profiles_csv*
- */db/{database}/schemes/{scheme_id}/profiles/{profile_id}*
- */db/{database}/isolates*
- */db/{database}/isolates/{isolate_id}*
- */db/{database}/isolates/{isolate_id}/allele_designations*

- */db/{database}/isolates/{isolate_id}/allele_designations/{locus}*
- */db/{database}/isolates/{isolate_id}/allele_ids*
- */db/{database}/isolates/{isolate_id}/schemes/{scheme_id}/allele_designations*
- */db/{database}/isolates/{isolate_id}/schemes/{scheme_id}/allele_ids*
- */db/{database}/isolates/{isolate_id}/contigs*
- */db/{database}/isolates/{isolate_id}/contigs_fasta*
- */db/{database}/contigs/{contig_id}*
- */db/{database}/fields*
- */db/{database}/users/{user_id}*

15.2.1 / or /db

Lists database resources available using the API.

Supported methods: GET, POST

Required query parameters: None

Optional query parameters: None

Example request URI: <http://rest.pubmlst.org/>

Response: List of resource groupings (ordered by name). Groups may consist of paired databases for sequence definitions and isolate data, or any set of related resources. Each group contains:

- *name* [string] - short name (usually a single word)
- *description* [string] - fuller description
- *databases* [array] - list of database objects, each consists of three key/value pairs:
 - *name* [string] - name of database config
 - *description* [string] - short description of resource
 - *href* [string] - URI to access resource

15.2.2 /db/{database}

Lists resources available for specified database configuration. These will vary depending on whether the resource is an isolate or a sequence definition database.

Supported methods: GET, POST

Required query parameter: {database} - Database configuration name [string]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates

Response: Object containing a subset of the following key/value pairs:

- *fields* [string] - URI to isolate provenance field information
- *isolates* [string] - URI to isolate records
- *schemes* [string] - URI to list of schemes

- *loci* [string] - URI to list of loci
- records [integer] - count of available records

15.2.3 /db/{database}/loci

Lists loci defined within specified database configuration.

Supported methods: GET, POST

Required query parameter: {database} - Database configuration name [string]

Optional parameters:

- page [integer]
- page_size [integer] - Set very large page size to return all results in one go.

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci

Response: Object containing:

- 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

15.2.4 /db/{database}/loci/{locus}

Provides information about a locus, including links to allele sequences (in seqdef databases).

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {locus} - Locus name [string]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci/abcZ

Response: Object containing a subset of the following key/value pairs:

- id [string] - locus name
- data_type [string] - 'DNA' or 'peptide'
- allele_id_format [string] - 'integer' or 'text'
- allele_id_regex [string] - regular expression constraining allele ids
- common_name [string]
- aliases [array] - list of alternative names of the locus
- length_varies [boolean]

- length [integer] - length if alleles are of a fixed length
- coding_sequence [boolean]
- orf [integer] - 1-6
- **schemes [array] - list of scheme objects, each consisting of:**
 - scheme [string] - URI to scheme information
 - description [string]
- min_length [integer] (seqdef databases) - minimum length for variable length loci
- max_length [integer] (seqdef databases) - maximum length for variable length loci
- alleles [string] (seqdef databases) - *URI to list of allele records*
- alleles_fasta [string] (seqdef databases) - *URI to FASTA file of all alleles of locus*
- curators [array] (seqdef databases) - list of URIs to user records of curators of the locus
- publications [array] (seqdef databases) - list of PubMed id numbers of papers describing the locus
- full_name [string] (seqdef databases)
- product [string] (seqdef databases)
- description [string] (seqdef databases)
- extended_attributes [array] (seqdef databases) - list of extended attribute objects. Each consists of a subset of the following fields:
 - field [string] - field name
 - value_format [string] - 'integer', 'text', or 'boolean'
 - value_regex [string] - regular expression constraining value
 - description [string] - description of field
 - length [integer] - maximum length of field
 - required [boolean]
 - allowed_values [array] - list of allowed values
- genome_position [integer] (isolate databases)

15.2.5 /db/{database}/loci/{locus}/alleles

Lists alleles defined for specific locus.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {locus} - Locus name [string]

Optional parameters:

- page [integer]
- page_size [integer] - Set very large page size to return all results in one go.

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci/abcZ/alleles

Response: Object containing:

- 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

15.2.6 /db/{database}/loci/{locus}/alleles_fasta

Provides all alleles defined for a locus in FASTA format.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {locus} - Locus name [string]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci/abcZ/alleles_fasta

Response: FASTA format file of alleles sequences

15.2.7 /db/{database}/loci/{locus}/alleles/{allele_id}

Provides information about an allele including its sequence.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {locus} - Locus name [string]
- {allele_id} - Allele identifier [string]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci/abcZ/alleles/5

Response: Object containing the following key/value pairs:

- locus [string] - *URI to locus description*
- allele_id [string] - allele identifier
- sequence [string] - sequence
- status [string] - either 'Sanger trace checked', 'WGS: manual extract', 'WGS: automated extract', or 'unchecked'
- sender [string] - *URI to user details* of sender

- curator [string] - *URI to user details* of curator
- date_entered [string] - record creation date (ISO 8601 format)
- datestamp [string] - last updated date (ISO 8601 format)

15.2.8 /db/{database}/schemes

Lists schemes defined within specified database configuration.

Supported methods: GET, POST

Required query parameter: {database} - Database configuration name [string]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes

Response: List of scheme objects, each containing:

- scheme [string] - URI to scheme information
- description [string]

15.2.9 /db/{database}/schemes/{scheme_id}

Provides information about a scheme, including links to allelic profiles (in seqdef databases, if appropriate).

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {scheme_id} - Scheme id number [integer]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1

Response: Object containing a subset of the following key/value pairs:

- id [integer]
- description [string]
- locus_count [integer] - number of loci belonging to scheme
- loci [array] - list of *URIs to locus descriptions*
- has_primary_key_field [boolean]
- fields [array] - list of *URIs to scheme field descriptions*
- primary_key_field [string] - *URI to primary key field description*
- profile_count [integer] - number of defined profiles (only for schemes with primary keys defined - only seqdef databases)
- profiles [array] - URI to list of profile definitions (only seqdef databases)
- profiles_csv [string] - URI to tab-delimited file of all scheme profiles

15.2.10 /db/{database}/schemes/{scheme_id}/fields/{field}

Provides information about scheme fields.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {scheme_id} - Scheme id number [integer]
- {field} - Field name [string]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/fields/ST

Response: Object containing the following key/value pairs:

- field [string] - field name
- type [string] - data type of field (integer or text)
- primary_key [boolean] - true if field is the scheme primary key

15.2.11 /db/{database}/schemes/{scheme_id}/profiles

Lists allelic profiles defined for a specific scheme.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {scheme_id} - Scheme id [integer]

Optional parameters:

- page [integer]
- page_size [integer] - Set very large page size to return all results in one go.

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/profiles

Response: Object containing:

- 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

15.2.12 /db/{database}/schemes/{scheme_id}/profiles_csv

Provides all profiles defined for a scheme in CSV (tab-delimited) format.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {scheme_id} - Scheme id [integer]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/profiles_csv

Response: Tab-delimited text file of allelic profiles

15.2.13 /db/{database}/schemes/{scheme_id}/profiles/{profile_id}

Provides information about a specific allelic profile defined for a scheme.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {scheme_id} - Scheme id [integer]
- {profile_id} - Profile id [string/integer]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/profiles/11

Response: Object containing the following key/value pairs:

- *primary_key_term* [string/integer] - The field name is the primary key, e.g. ST. The value is the primary key value (primary_id used as an argument).
- *alleles* [object] - *list of URIs to allele descriptions*
- *other_scheme_fields* [string/integer] - Each scheme field will have its own value if defined. The field name is the name of the field.
- *sender* [string] - *URI to user details* of sender
- *curator* [string] - *URI to user details* of curator
- *date_entered* [string] - record creation date (ISO 8601 format)
- *timestamp* [string] - last updated date (ISO 8601 format)

15.2.14 /db/{database}/isolates

Provides list of isolate records.

Supported methods: GET, POST

Required query parameter: {database} - Database configuration name [string]

Optional parameters:

- *page* [integer]

- `page_size` [integer] - Set very large page size to return all results in one go.

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates

Response: Object containing:

- `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

15.2.15 /db/{database}/isolates/{isolate_id}

Provides information about an isolate.

Supported methods: GET, POST

Required query parameters:

- `{database}` - Database configuration name [string]
- `{isolate_id}` - Isolate identifier [integer]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1

Response: Object containing some or all of the following key/value pairs:

- `provenance` [object] - set of key/value pairs. Keys are defined by calling the */fields route*. The fields will vary by database but will always contain the following:
 - `id` [integer]
 - `sender` [string] - *URI to user details* of sender
 - `curator` [string] - *URI to user details* of curator
 - `date_entered` [string] - record creation date (ISO 8601 format)
 - `timestamp` [string] - last updated date (ISO 8601 format)
- `publications` [array] (seqdef databases) - list of PubMed id numbers of papers that refer to the isolate
- `sequence_bin` [object] - consists of the following key/value pairs:
 - `contigs_fasta` [string] - *URI to FASTA file containing all the contigs belonging to this isolate*
 - `contigs` [string] - *URI to list of contig records*
 - `contig_count` [integer] - number of contigs
 - `total_length` [integer] - total length of contigs
- `allele_designations` [object] - consists of the following key/value pairs:
 - `allele_ids` - *URI to list of all allele_id values* defined for the isolate
 - `designation_count` - number of allele designations defined for the isolate
 - `full_designations` - *URI to list of full allele designation records*

- schemes [array] - list of scheme objects, each containing the following:
 - description [string] - description of scheme
 - loci_designated_count [integer] - number of loci within scheme that have an allele designated for this isolate.
 - allele_ids [string] - *URI to list of all allele_id values defined for this scheme* for this isolate
 - full_designations [string] - *URI to list of full allele designation records* for this isolate
 - fields [object] - consisting of key/value pairs where the key is the name of each scheme field

15.2.16 /db/{database}/isolates/{isolate_id}/allele_designations

Provides a list of full allele designation records for the specified isolate.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {isolate_id} - Isolate identifier [integer]

Optional parameters:

- page [integer]
- page_size [integer] - Set very large page size to return all results in one go.

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/allele_designations

Response: Object containing:

- 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

15.2.17 /db/{database}/isolates/{isolate_id}/allele_designations/{locus}

Provides a full allele designation record.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {isolate_id} - Isolate identifier [integer]
- {locus} - Locus name [string]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/allele_designations/BACT000065

Response: List of allele_designation objects (there may be multiple designations for the same locus), each containing:

- locus [string] - *URI to locus description*
- allele_id [string]
- method [string] - either 'manual' or 'automatic'
- status [string] - either 'confirmed' or 'provisional'
- comments [string]
- sender [string] - *URI to user details* of sender
- curator [string] - *URI to user details* of curator
- timestamp [string] - last updated date (ISO 8601 format)

15.2.18 /db/{database}/isolates/{isolate_id}/allele_ids

Returns array of allele identifiers for isolate.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {isolate_id} - Isolate identifier [integer]

Optional parameters:

- page [integer]
- page_size [integer] - Set very large page size to return all results in one go.

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/allele_ids

Response: Object containing:

- 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

15.2.19 /db/{database}/isolates/{isolate_id}/schemes/{scheme_id}/allele_designations

Provides a list of full allele designation records for loci belonging to the specified scheme and isolate.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]

- {isolate_id} - Isolate identifier [integer]
- {scheme_id} - Scheme identifier [integer]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/schemes/1/allele_designations

Response: Array containing *allele designation objects* for each locus in the specified scheme that has been designated.

15.2.20 /db/{database}/isolates/{isolate_id}/schemes/{scheme_id}/allele_ids

Provides a list of allele identifiers for loci belonging to the specified scheme and isolate.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {isolate_id} - Isolate identifier [integer]
- {scheme_id} - Scheme identifier [integer]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/schemes/1/allele_ids

Response: Array containing allele id objects for each locus in the specified scheme that has been designated. Each allele_id object contains a key which is the name of the locus with a value that may be either a string, integer or array of strings or integers (required where there are multiple designations for a locus). The data type depends on the allele_id_format set for the specific locus.

15.2.21 /db/{database}/isolates/{isolate_id}/contigs

Returns a list of contig records for the specified isolate.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {isolate_id} - Isolate identifier [integer]

Optional parameters:

- page [integer]
- page_size [integer] - Set very large page size to return all results in one go.

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/contigs

Response: Object containing:

- 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

15.2.22 /db/{database}/isolates/{isolate_id}/contigs_fasta

Provides all contigs associated with an isolate record in FASTA format.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {isolate_id} - Isolate identifier [integer]

Optional parameter:

- header [string] - either 'original_designation' or 'id' (default is 'id'). This selects whether the FASTA header lines contain the originally uploaded FASTA headers or the sequence bin id numbers.

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/contigs_fasta?header=original_designation

Response: FASTA format file of isolate contig sequences

15.2.23 /db/{database}/contigs/{contig_id}

Provides a record of a specified contig.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {contig_id} - Contig identifier [integer]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/contigs/180062

Response: Contig object consisting of the following key/value pairs:

- id [integer] - contig identifier
- isolate_id [integer] - isolate identifier
- sequence [string] - contig sequence
- length [integer] - length of contig sequence
- method [string] - sequencing method
- sender [string] - *URI to user details* of sender
- curator [string] - *URI to user details* of curator
- date_entered [string] - record creation date (ISO 8601 format)
- timestamp [string] - last updated date (ISO 8601 format)

15.2.24 /db/{database}/fields

Provides a list of isolate provenance field descriptions.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/fields

Response: Array of field objects, each consisting of some or all of the following key/value pairs:

- name [string] - name of field
- type [string] - data type (int, text, date, float)
- length [integer] - maximum length of field
- required [boolean] - true if field value is required
- min [integer] - minimum value for integer values
- max [integer] - maximum value for integer values
- regex [string] - regular expression that constrains the allowed value of the field
- comments [string]
- allowed values [array] - list of allowed values for the field [string]

15.2.25 /db/{database}/users/{user_id}

Provides information about data senders and curators.

Supported methods: GET, POST

Required query parameters:

- {database} - Database configuration name [string]
- {user_id} - User id number [integer]

Optional parameters: None

Example request URI: http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/users/2

Response: Object containing the following key/value pairs:

- id [integer] - user id number
- first_name [string]
- surname [string]
- affiliation [string] - institutional affiliation
- email [string] - E-mail address

15.3 Authentication

Protected resources, i.e. those requiring a user to log in, can be accessed via the API using OAuth (1.0A) authentication (see [IETF RFC5849](#) for details). Third-party client software has to be registered with the BIGSdb site before they can access authenticated resources. The overall three-legged flow works as follows:

1. *Developer signs up* and gets a consumer key and consumer secret specific to their application.
2. Application *gets a request token* and directs user to authorization page on BIGSdb.
3. BIGSdb *asks user for authorization* for application to access specific resource using their credentials. A verifier code is provided.
4. The application exchanges the request token and OAuth verifier code for an *access token and secret* (these do not expire but may be revoked by the user or site admin).
5. Application uses access token/secret to *request session token* (this is valid for 12 hours).
6. All calls to *access protected resources* are signed using the session token/secret and consumer key/secret.

It is recommended that application developers use an OAuth library to generate and sign requests.

15.3.1 Developer sign up to get a consumer key

Application developers should apply to the site administrator of the site running BIGSdb. The administrator can *generate a key and secret* using a script - both of these will need to be used by the application to sign requests.

The client id is usually a 24 character alphanumeric string. The secret is usually a 42 character alphanumeric (including punctuation) string, e.g.

- **client_id:** efKXmqp2D0EB1MBkZaGC2lPf
- **client_secret:** F\$M)_+fQ2AFFB2YBDfF9fpHF^qSWJdmmN%L4Fxf5Gur3

15.3.2 Getting a request token

- **Relative URL:** /db/{database}/oauth/get_request_token
- **Supported methods:** GET, POST

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` ('POST')
- `oauth_request_url` (request URL)
- `oauth_signature_method` ('HMAC-SHA1')
- `oauth_signature`
- `oauth_timestamp` (UNIX timestamp - seconds since Jan 1 1970) - this must be within 600 seconds of the current time.
- `oauth_callback` ('oob' for desktop applications)
- `oauth_nonce` (random string)
- `oauth_version` ('1.0')

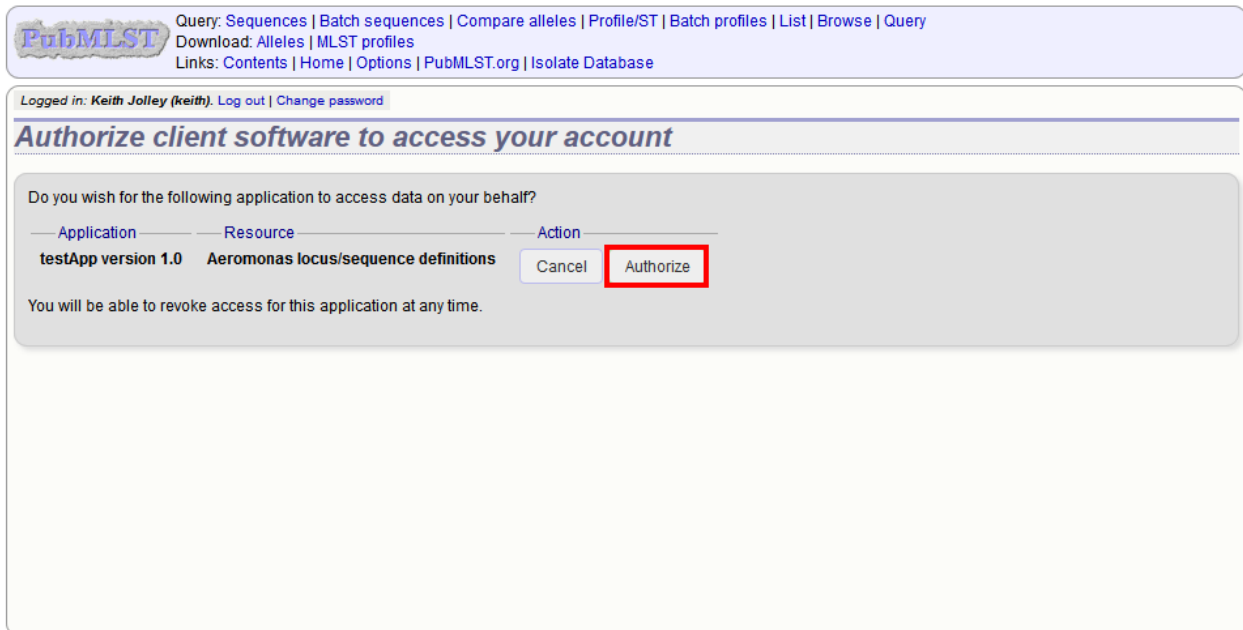
If the application has been registered and has been granted permission to access the specific resource, a JSON response will be returned containing the following parameters:

- **oauth_token**
 - This is the request token. It is usually a 32 character alphanumeric string.
 - e.g. fKFm0WNhCfbEX8zQm6qhDA8K23FOWDGE
- **oauth_token_secret**
 - This is the secret associated with the request token. It is usually a 32 character alphanumeric string.
 - e.g. aZ0fncP7i5w5jlebdK5zyQ4vrRRVcdnv
- **oauth_callback_confirmed**
 - This parameter is always set to true.

15.3.3 Getting user authorization

Once a request token has been obtained, this can be used by the end user to grant permission to access a specific resource to the application. The application should direct the user to the client authorization page (authorizeClient) specific to a database within BIGSdb, e.g. http://pubmlst.org/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_neisseria_seqdef&page=authorizeClient&oauth_token=fKFm0WNhCfbEX8zQm6qhDA8K23FOWDGE

The user will be asked if they wish to grant access to the application on their behalf:



PubMLST Query: [Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
Download: [Alleles](#) | [MLST profiles](#)
Links: [Contents](#) | [Home](#) | [Options](#) | [PubMLST.org](#) | [Isolate Database](#)

Logged in: **Keith Jolley (keith)** | [Log out](#) | [Change password](#)


Authorize client software to access your account

Do you wish for the following application to access data on your behalf?

| Application | Resource | Action |
|---------------------|--------------------------------------|--|
| testApp version 1.0 | Aeromonas locus/sequence definitions | <input type="button" value="Cancel"/> <input type="button" value="Authorize"/> |

You will be able to revoke access for this application at any time.

If they authorize the access, they will be presented with a verifier code. This should be entered in to the client application which will use this together with the request token to request an access token.


[Query: Sequences](#) | [Batch sequences](#) | [Compare alleles](#) | [Profile/ST](#) | [Batch profiles](#) | [List](#) | [Browse](#) | [Query](#)
[Download: Alleles](#) | [MLST profiles](#)
[Links: Contents](#) | [Home](#) | [Options](#) | [PubMLST.org](#) | [Isolate Database](#)

Logged in: [Keith Jolley \(keith\)](#). [Log out](#) | [Change password](#)

Authorize client software to access your account

You have authorized **testApp version 1.0** to access **Aeromonas locus/sequence definitions** on your behalf.

Enter the following verification code when asked by testApp.

Verification code: 2qaMxNrP

This code is valid for 60 minutes.

The verifier code is valid for 60 minutes.

15.3.4 Getting an access token

- **Relative URL:** /db/{database}/oauth/get_access_token
- **Supported methods:** GET, POST

The application uses the request token, verifier code and its consumer key to obtain an access token. The access token does not expire but can be revoked by both the end user or the site administrator. The request needs to contain the following parameters and to be signed using the consumer secret and request token secret:

- `oauth_consumer_key`
- `oauth_request_method` ('POST')
- `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. SDrC74ZVl5SYSqY8lWZqrRxnyDnNGVFO
- **oauth_token_secret**
 - This is the secret associated with the access token. It is usually a 32 character alphanumeric string.
 - e.g. tYl2SPzgiO02lRVzW4JR1ez6Vvm4gVyv

15.3.5 Getting a session token

- **Relative URL:** /db/{database}/oauth/get_session_token
- **Supported methods:** GET, POST

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** ('POST')
- **oauth_request_url** (request URL)
- **oauth_signature_method** ('HMAC-SHA1')
- **oauth_signature**
- **oauth_token** (access token)
- **oauth_timestamp** (UNIX timestamp - seconds since Jan 1 1970) - this must be within 600 seconds of the current time.
- **oauth_nonce** (random string)
- **oauth_version** ('1.0')

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

- **oauth_token**
 - This is the session token. It is usually a 32 character alphanumeric string.
 - e.g. H8CjIS8Ikq6hwCUqUfF1l4pTaCYl8Ljw
- **oauth_token_secret**
 - This is the secret associated with the session token. It is usually a 32 character alphanumeric string.
 - e.g. RfponbaNPO7tkZ2miHFISk0pMndePNfJ

15.3.6 Accessing protected resources

- **Supported methods:** GET, POST

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` ('POST')
- `oauth_request_url` (request URL)
- `oauth_signature_method` ('HMAC-SHA1')
- `oauth_signature`
- `oauth_token` (session token)
- `oauth_timestamp` (UNIX timestamp - seconds since Jan 1 1970) - this must be within 600 seconds of the current time.
- `oauth_nonce` (random string)
- `oauth_version` ('1.0')

Frequently asked questions (FAQs)

16.1 General

1. What is the minimum specification of hardware required to run BIGSdb?

The software will run on very modest hardware - a number of PubMLST mirrors have been set up on virtual machines with 1 processor core and 4 GB RAM. This should be considered an absolute minimum specification though. For an installation with only local users, the following minimum is recommended:

- 4 processor cores
- 16 GB RAM
- 50 GB partition for temporary files
- 100 GB partition for databases

As usual, the more RAM that is available the better. Ideally you would want enough RAM that the whole database(s) can reside in memory (an approximation is roughly twice the total size of your contigs), although this is not absolutely required.

Offline jobs, such as *Genome Comparator* will use a processor core each, so if you want to run multiple jobs in parallel then you may want more cores (and memory). Tagging of new genomes using the offline *autotagger* can be run in multi-threaded mode so the more cores available the faster this will be.

As a comparison, the PubMLST site is run on two machines - separate web and database servers. All offline jobs and tagging of genomes is performed on the database server. These have the following specification:

- web server: 16 cores, 64GB RAM
- database server: 64 cores, 1TB RAM, 3TB RAID 10 local storage

16.2 Installation

1. BIGSdb is accumulating files in various temp directories - is this normal and how do I clean them out?

See: *Periodically delete temporary files*.

2. BIGSdb is complaining of an invalid script path - what does this mean?

In your database config.xml file system tag are two attributes - script_path_includes and curate_path_includes. These contain regexes that the web url to your script (bigsdbs.pl and bigscurate.pl

respectively) must match. This prevents somebody from accessing a private database using an instance of bigsdb.pl that is not in a protected directory if you're using apache authentication.

So, if you access the script from <http://localhost/cgi-bin/bigsdb/bigsdb.pl> then you can set script_path_includes to something like "/bigsdb/" (which is the default), or "/cgi-bin/" or just "/" if you don't care about this check.

16.3 Administration

1. How can I make some isolates public but not others?

The easiest way to do this is to set up two or more separate configuration directories that refer to the database. The URLs to access these will differ by the value of the 'db' attribute, which refers to the name of the configuration directory (in /etc/bigsdb/dbases/). The database view accessed by each of these configurations can be different as can the access restrictions.

Example:

We have a database 'bigsdb_test' that contains data, only some of which we wish to make publicly available. The isolates to make public are all members of a project. First we can make a view of the isolates table that contains only isolates within this project.

For isolates in project id 3, create a database view by logging in to psql as the postgresql user. We will name this view 'public':

```
sudo su postgres
psql bigsdb_test

CREATE VIEW public AS SELECT * FROM isolates WHERE id IN (SELECT isolate_id
FROM project_members WHERE project_id=3);
GRANT SELECT ON public TO apache;
```

Create a private configuration that can access everything in the database in /etc/bigsdb/dbases/test_private. This will be accessible from http://IP_ADDRESS/cgi-bin/bigsdb/bigsdb.pl?db=test_private.

The important attributes to set in the system tag of the config.xml file in this directory are::

```
view="isolates"
read_access="authenticated_users"
```

This means that anyone with an account can log in and view all the isolates (because the view is set to the isolates table).

Now create a public configuration in /etc/bigsdb/dbases/test_public. This will be accessible from http://IP_ADDRESS/cgi-bin/bigsdb/bigsdb.pl?db=test_public. It is better to create a symlink to the private config.xml and then override the attributes that are different. So create a symlink to the private config file:

```
cd /etc/bigsdb/dbases/test_public
sudo ln -s ../test_private/config.xml .
```

You can now override the view and access settings. Within /etc/bigsdb/dbases/test_public, create a file called system.overrides and add the following:

```
view="public"
read_access="public"
```

See also *Restricting particular configurations to specific user accounts*.

17.1 Query operators

Various query forms have operators for use with field values. Available operators are:

- =
 - Exact match (case-insensitive).
- contains
 - Match to a partial string (case-insensitive), e.g. searching for clonal complex ‘contains’ st-11 would return all STs belonging to the ST-11 complex.
- starts with
 - Match to values that start with the search term (case-insensitive).
- ends with
 - Match to values that end with the search term (case-sensitive).
- >
 - Greater than the search term.
- <
 - Less than the search term.
- NOT
 - Match to values that do not equal the search term (case-insensitive).
- NOT contain
 - Match to values that do not contain the search term (case-insensitive).

17.2 Sequence tag flags

Sequences tagged in the sequence bin can have features indicated by specific flags. The presence of these flags can be queried. These are a superset of *flags available for allele sequences*. Available flags are:

- ambiguous read
 - Genome sequence contains ambiguous nucleotides in coding sequence.

- apparent misassembly
 - Sequence has a region of very high identity to existing allele in one region but looks completely different in another.
- atypical
 - Catch-all term for a sequence that is unusual compared to other alleles of locus.
- contains IS element
 - Coding sequence is interrupted by insertion sequence.
- downstream fusion
 - No stop codon present resulting in translation continuing.
- frameshift
 - Frameshift in sequence relative to other alleles, not resulting in internal stop codon.
- internal stop codon
 - Frameshift in sequence relative to other alleles, resulting in internal stop codon.
- no start codon
 - No apparent start codon in immediate vicinity of usual start.
- phase variable: off
 - Coding sequence has a homopolymeric run with a frameshift resulting in a stop codon preventing complete translation.
- truncated
 - Coding sequence is unusually short resulting in a truncated protein (not the same as running off the end of a contig).
- upstream fusion
 - No apparent start codon in immediate vicinity of usual start, likely due to a gene fusion (sequence is transcribed together with upstream coding sequence).

17.3 Allele sequence flags

Sequences can be flagged with specific attributes - these are searchable when doing a sequence attribute query. These are a subset of *flags available for tagged sequences*. These are mainly for use with whole genome MLST type data. Multiple flags can be selected by Ctrl-clicking the list. Available flags are:

- atypical
 - Catch-all term for a sequence that is unusual compared to other alleles of locus.
- contains IS element
 - Coding sequence is interrupted by insertion sequence.
- downstream fusion
 - No stop codon present resulting in translation continuing.
- frameshift
 - Frameshift in sequence relative to other alleles, not resulting in internal stop codon.

- internal stop codon
 - Frameshift in sequence relative to other alleles, resulting in internal stop codon.
- no start codon
 - No apparent start codon in immediate vicinity of usual start.
- phase variable: off
 - Coding sequence has a homopolymeric run with a frameshift resulting in a stop codon preventing complete translation.
- truncated
 - Coding sequence is unusually short resulting in a truncated protein (not the same as running off the end of a contig).
- upstream fusion
 - No apparent start codon in immediate vicinity of usual start, likely due to a gene fusion (sequence is transcribed together with upstream coding sequence).

Database schema

- Sequence definition database
- Isolate database

A

access

- control lists, 32
- restricting, 32

adding

- locus, 40, 43, 44, 50, 74
- MLST scheme, 73
- schemes, 55

allele

- breakdown, 241

allele definition

- records, 185

allele designations

- query, 207
- status, 208

allele sequence

- identify, 189

allele sequences

- adding, 100

alleles, 3

API authentication

- access token, 331
- accessing protected resources, 332
- consumer key, 329
- request token, 329
- session token, 332
- user authorization, 330

API resources

- /, 316
- /db, 316
- /db/{database}, 316
- /db/{database}/contigs/{contig_id}, 327
- /db/{database}/fields, 327
- /db/{database}/isolates, 322
- /db/{database}/isolates/{isolate_id}, 323
- /db/{database}/isolates/{isolate_id}/allele_designations, 324
- /db/{database}/isolates/{isolate_id}/allele_designations/{locus}, 324
- /db/{database}/isolates/{isolate_id}/allele_ids, 325

/db/{database}/isolates/{isolate_id}/contigs, 326

/db/{database}/isolates/{isolate_id}/contigs_fasta, 327

/db/{database}/isolates/{isolate_id}/schemes/{scheme_id}/allele_designations, 325

/db/{database}/isolates/{isolate_id}/schemes/{scheme_id}/allele_ids, 326

/db/{database}/loci, 317

/db/{database}/loci/{locus}, 317

/db/{database}/loci/{locus}/alleles, 318

/db/{database}/loci/{locus}/alleles/{allele_id}, 319

/db/{database}/loci/{locus}/alleles_fasta, 319

/db/{database}/schemes, 320

/db/{database}/schemes/{scheme_id}, 320

/db/{database}/schemes/{scheme_id}/fields/{field}, 320

/db/{database}/schemes/{scheme_id}/profiles, 321

/db/{database}/schemes/{scheme_id}/profiles/{profile_id}, 322

/db/{database}/schemes/{scheme_id}/profiles_csv, 321

/db/{database}/users/{user_id}, 328

auto allele definer, 171

- stop, 173

autotagger, 169

- stop, 173

B

BLAST, 257

breakdown

- allele, 241
- provenance field, 236
- scheme, 241
- sequence bin, 246
- two-field, 239

browse

- scheme profiles, 197

BURST, 261

C

caching

- schemes, 35
- client authorization
 - RESTful interface, 96
- client databases, 65
- codon usage, 265
 - Locus Explorer, 234
- composite fields, 85
- configuration
 - export, 94
- configuration settings
 - validation, 93

E

- export
 - configuration, 94
- extended attributes
 - locus, 52
 - provenance fields, 88

F

- filters, 211

G

- Genome Comparator, 248
- genome filtering, 76
 - in silico hybridization, 79
 - in silico PCR, 76
- groups
 - scheme, 61

H

- hosts
 - mapping, 35

I

- identify
 - allele sequence, 189
 - sequence type, 191
- in silico hybridization, 79
- in silico PCR, 76
- isolate
 - records, 181
- isolate record
 - options, 222

L

- loci, 3
 - modifying display, 227
- locus
 - adding, 40, 43, 44, 50, 74
 - copying existing record, 48
 - extended attributes, 52
- Locus Explorer, 231

- codon usage, 234
- polymorphic sites, 231
- translated sequences, 235
- locus positions
 - setting, 80

M

- main results table
 - options, 221
- mapping
 - hosts, 35
- materialized views, 36
- MLST, 3
- MLST scheme
 - adding, 73
- mod_perl, 35
- modifying display
 - loci, 227
 - schemes, 227

O

- offline curation
 - auto allele definer, 171
 - autotagger, 169
- options, 219
 - isolate record, 222
 - main results table, 221
 - provenance fields, 224
 - query, 225

P

- partitioning
 - sets, 36
- passwords
 - setting, 32
 - first user, 33
- performance
 - caching schemes, 35
 - materialized views, 36
 - mod_perl, 35
- permissions, 27
 - locus curation, 30
 - scheme curation, 30
- plugins
 - enabling, 33
- polymorphic sites
 - Locus Explorer, 231
- polymorphisms, 270
- profile
 - records, 187
- profiles, 4
- projects, 141
- provenance field
 - breakdown, 236

provenance fields
 extended attributes, 88
 options, 224

Q

query
 allele designations, 207
 options, 225
 scheme profiles, 199
 sequence tags, 209

R

records
 allele definition, 185
 isolate, 181
 profile, 187
 sequence bin, 187
 sequence tag, 186
RESTful interface
 client authorization, 96
rule-based queries, 71

S

scheme
 breakdown, 241
 groups, 61
scheme profiles
 browse, 197
 query, 199
schemes, 3
 adding, 55
 caching, 35
 modifying display, 227
sequence bin
 breakdown, 246
 records, 187
sequence similarity
 determining, 200
sequence tag
 records, 186
 status, 274
sequence tags, 4
 query, 209
sequence type
 identify, 191
sets, 4
 partitioning, 36
stop
 auto allele definer, 173
 autotagger, 173

T

translated sequences

Locus Explorer, 235
two-field
 breakdown, 239

U

unique combinations, 268
updates
 disabling, 34
user groups, 27
user types, 27
users
 adding, 99