

---

# **BIGSdb Documentation**

***Release 1.22.0***

**Keith Jolley**

**Mar 07, 2019**



<b>1</b>	<b>Concepts and terms</b>	<b>3</b>
1.1	BIGSdb	3
1.2	Loci	3
1.3	Alleles	3
1.4	Schemes	4
1.5	Profiles	4
1.6	Classification groups	4
1.7	Sequence tags	4
1.8	Sets	5
<b>2</b>	<b>BIGSdb dependencies</b>	<b>7</b>
2.1	Required packages	7
2.1.1	Linux packages	7
2.1.2	Perl modules	7
2.1.3	Optional packages	8
<b>3</b>	<b>Installation and configuration of BIGSdb</b>	<b>9</b>
3.1	Software installation	9
3.2	Configuring PostgreSQL	10
3.2.1	pg_hba.conf	10
3.2.2	pg_ident.conf	11
3.3	Setting global connection parameters	11
3.4	Site-specific configuration	12
3.5	Setting up the offline job manager	12
3.6	Setting up the submission system	13
3.7	Setting up a site-wide user database	14
3.8	Periodically delete temporary files	15
3.9	Prevent preference database getting too large	15
3.10	Log file rotation	16
3.11	Upgrading BIGSdb	16
3.12	Running the BIGSdb RESTful interface	16
3.12.1	Proxying the API to use a standard web port	19
<b>4</b>	<b>Database setup</b>	<b>21</b>
4.1	Creating databases	21
4.2	Database-specific configuration	22
4.3	XML configuration attributes used in config.xml	23

4.3.1	Isolate database XML attributes	23
4.3.1.1	Special values	30
4.3.2	Sequence definition database XML attributes	30
4.4	Over-riding global defaults set in bigsdb.conf	33
4.5	Over-riding values set in config.xml	34
4.6	User authentication	34
4.6.1	Apache authentication	34
4.6.2	Built-in authentication	35
4.7	Setting up the admin user	35
4.8	Updating PubMed citations	35
4.9	Configuring access to remote contigs	36
4.9.1	Setting up authentication	36
4.9.2	Processing remote contigs	37
<b>5</b>	<b>Administrator's guide</b>	<b>39</b>
5.1	Types of user	39
5.2	User groups	39
5.3	Curator permissions	40
5.4	Locus and scheme permissions (sequence definition database)	42
5.5	Controlling access	44
5.5.1	Restricting particular configurations to specific user accounts	44
5.6	Setting user passwords	45
5.7	Setting the first user password	45
5.8	Enabling plugins	45
5.9	Temporarily disabling database updates	46
5.9.1	Global	46
5.9.2	Database-specific	47
5.10	Host mapping	47
5.11	Improving performance	47
5.11.1	Use mod_perl	47
5.11.2	Cache scheme definitions within an isolate database	47
5.11.3	Use a ramdisk for the secure temporary directory	49
5.12	Dataset partitioning	49
5.12.1	Sets	49
5.12.2	Configuration of sets	49
5.12.3	Set metadata	50
5.12.4	Set views	51
5.12.5	Using only defined sets	52
5.13	Setting a site-wide users database	52
5.14	Adding new loci	53
5.14.1	Sequence definition databases	53
5.14.1.1	Single locus	53
5.14.1.2	Batch adding	55
5.14.2	Isolate databases	57
5.14.2.1	Single locus	57
5.14.2.2	Batch adding	62
5.15	Defining locus extended attributes	63
5.16	Defining schemes	65
5.16.1	Sequence definition databases	65
5.16.2	Isolate databases	69
5.17	Organizing schemes into hierarchical groups	71
5.18	Setting up client databases	74
5.18.1	Look up isolates with given allele	76
5.18.2	Look up isolates with a given scheme primary key	77



5.18.3	Look up specific isolate database fields linked to a given allele . . . . .	78
5.19	Rule-based sequence queries . . . . .	80
5.19.1	Example rule code . . . . .	80
5.19.1.1	Rule files . . . . .	82
5.19.1.2	Linking to the rule query . . . . .	82
5.19.1.3	Adding descriptive text . . . . .	82
5.20	Workflow for setting up a MLST scheme . . . . .	82
5.21	Automated assignment of scheme profiles . . . . .	83
5.22	Scheme profile clustering - setting up classification schemes . . . . .	84
5.22.1	Defining classification scheme in sequence definition database . . . . .	84
5.22.2	Defining classification scheme in isolate database . . . . .	86
5.22.3	Clustering . . . . .	87
5.23	Defining new loci based on annotated reference genome . . . . .	89
5.24	Genome filtering . . . . .	91
5.24.1	Filtering by <i>in silico</i> PCR . . . . .	91
5.24.2	Filtering by <i>in silico</i> hybridization . . . . .	94
5.25	Setting locus genome positions . . . . .	96
5.26	Defining composite fields . . . . .	98
5.27	Extended provenance attributes (lookup tables) . . . . .	101
5.28	Sequence bin attributes . . . . .	103
5.29	Checking external database configuration settings . . . . .	105
5.30	Exporting table configurations . . . . .	106
5.31	Authorizing third-party client software to access authenticated resources . . . . .	108
<b>6</b>	<b>Curator's guide</b>	<b>111</b>
6.1	Adding new sender details . . . . .	111
6.2	Adding new allele sequence definitions . . . . .	112
6.2.1	Single allele . . . . .	112
6.2.2	Batch adding multiple alleles . . . . .	115
6.2.2.1	Upload using a spreadsheet . . . . .	115
6.2.2.2	Upload using a FASTA file . . . . .	117
6.3	Updating and deleting allele sequence definitions . . . . .	119
6.4	Retiring allele identifiers . . . . .	122
6.5	Updating locus descriptions . . . . .	123
6.6	Adding new scheme profile definitions . . . . .	127
6.7	Updating and deleting scheme profile definitions . . . . .	131
6.8	Retiring scheme profile definitions . . . . .	132
6.9	Adding isolate records . . . . .	133
6.10	Updating and deleting single isolate records . . . . .	137
6.11	Batch updating multiple isolate records . . . . .	143
6.12	Deleting multiple isolate records . . . . .	145
6.13	Retiring isolate identifiers . . . . .	146
6.14	Setting alternative names for isolates (aliases) . . . . .	147
6.15	Linking isolate records to publications . . . . .	148
6.16	Uploading sequence contigs linked to an isolate record . . . . .	150
6.16.1	Select isolate from drop-down list . . . . .	150
6.16.2	Select from isolate query . . . . .	152
6.16.3	Upload options . . . . .	153
6.17	Batch uploading sequence contigs linked to multiple isolate records . . . . .	154
6.18	Linking remote contigs to isolate records . . . . .	157
6.19	Automated web-based sequence tagging . . . . .	159
6.20	Projects . . . . .	162
6.20.1	Creating the project . . . . .	162
6.20.2	Explicitly adding isolates to a project . . . . .	163

6.21	Isolate record versioning . . . . .	166
<b>7</b>	<b>Curating submitted data</b>	<b>169</b>
7.1	Alleles . . . . .	170
7.1.1	Individual allele curation . . . . .	171
7.1.2	Batch allele curation . . . . .	173
7.1.3	Rejecting sequences . . . . .	176
7.1.4	Requesting additional information . . . . .	176
7.1.5	Closing the submission . . . . .	176
7.2	Profiles . . . . .	177
7.2.1	Individual profile curation . . . . .	178
7.2.2	Batch profile curation . . . . .	181
7.2.3	Rejecting profiles . . . . .	183
7.2.4	Requesting additional information . . . . .	184
7.2.5	Closing the submission . . . . .	184
7.3	Isolates . . . . .	184
7.3.1	Requesting additional information . . . . .	187
7.3.2	Closing the submission . . . . .	187
<b>8</b>	<b>Offline curation tools</b>	<b>189</b>
8.1	Automated offline sequence tagging . . . . .	189
8.2	Defining exemplar alleles . . . . .	192
8.3	Automated offline allele definition . . . . .	193
8.4	Cleanly interrupting offline curation . . . . .	195
8.5	Uploading contigs from the command line . . . . .	195
<b>9</b>	<b>Definition downloads</b>	<b>197</b>
9.1	Allele sequence definitions . . . . .	197
9.1.1	Scheme tree . . . . .	198
9.1.2	Alphabetical list . . . . .	200
9.1.3	All loci by scheme . . . . .	201
9.1.4	Download locus table . . . . .	201
9.2	Scheme profile definitions . . . . .	202
<b>10</b>	<b>Data records</b>	<b>205</b>
10.1	Isolate records . . . . .	205
10.1.1	Projects . . . . .	207
10.1.2	Provenance metadata . . . . .	207
10.1.3	Publications . . . . .	208
10.1.4	Sequence bin summary . . . . .	209
10.1.5	Scheme and locus data . . . . .	209
10.2	Allele definition records . . . . .	210
10.3	Sequence tag records . . . . .	211
10.4	Profile records . . . . .	211
10.5	Sequence bin records . . . . .	212
<b>11</b>	<b>Querying data</b>	<b>213</b>
11.1	Querying sequences to determine allele identity . . . . .	213
11.1.1	Querying whole genome data . . . . .	215
11.2	Querying multiple sequences to identify allele identities . . . . .	216
11.3	Searching for specific allele definitions . . . . .	218
11.3.1	General (all loci) sequence attribute search . . . . .	218
11.3.2	Locus-specific sequence attribute search . . . . .	222
11.4	Browsing scheme profile definitions . . . . .	225
11.5	Querying scheme profile definitions . . . . .	226

11.6	Identifying allelic profile definitions . . . . .	230
11.7	Batch profile queries . . . . .	234
11.8	Investigating allele differences . . . . .	236
11.8.1	Sequence similarity . . . . .	236
11.8.2	Sequence comparison . . . . .	237
11.9	Browsing isolate data . . . . .	239
11.10	Querying isolate data . . . . .	241
11.10.1	Query by allele designation/scheme field . . . . .	244
11.10.2	Query by allele designation count . . . . .	245
11.10.3	Query by allele designation status . . . . .	247
11.10.4	Query by sequence tag count . . . . .	248
11.10.5	Query by tagged sequence status . . . . .	249
11.10.6	Query by list of attributes . . . . .	251
11.10.7	Query filters . . . . .	252
11.11	Querying by allelic profile . . . . .	253
11.12	Retrieving isolates by linked publication . . . . .	256
11.13	User-configurable options . . . . .	259
11.13.1	General options . . . . .	260
11.13.2	Main results table . . . . .	261
11.13.3	Isolate record display . . . . .	262
11.13.4	Provenance field display . . . . .	264
11.13.5	Query filters . . . . .	265
11.13.6	Modifying locus and scheme display options . . . . .	267
<b>12</b>	<b>User projects</b>	<b>271</b>
12.1	Adding isolates to a user project following a query . . . . .	272
12.2	Adding or removing isolates belonging to a user project by editing a list . . . . .	273
12.3	Accessing project isolates . . . . .	274
12.4	Allowing other users to share your project . . . . .	275
12.5	Deleting a user project . . . . .	276
<b>13</b>	<b>Private records</b>	<b>279</b>
13.1	Uploading private records . . . . .	279
13.2	Modifying private records . . . . .	281
13.3	Sharing access to private records . . . . .	282
<b>14</b>	<b>Data analysis plugins</b>	<b>283</b>
14.1	Locus explorer . . . . .	283
14.1.1	Polymorphic site analysis . . . . .	284
14.1.2	Codon usage . . . . .	287
14.1.3	Aligned translations . . . . .	288
14.2	Field breakdown . . . . .	289
14.2.1	Pie charts . . . . .	293
14.2.2	Bar charts . . . . .	295
14.2.3	Line charts . . . . .	296
14.2.4	Summary tables . . . . .	298
14.2.5	Exporting allele sequences . . . . .	299
14.3	Two field breakdown . . . . .	300
14.4	Sequence bin breakdown . . . . .	303
14.5	Genome comparator . . . . .	306
14.5.1	Analysis using defined loci . . . . .	307
14.5.2	Analysis using annotated reference genome . . . . .	310
14.5.3	Include in identifiers fieldset . . . . .	311
14.5.4	Reference genome fieldset . . . . .	311

14.5.5	Parameters/options fieldset . . . . .	311
14.5.6	Distance matrix calculation fieldset . . . . .	312
14.5.7	Alignments fieldset . . . . .	313
14.5.8	Core genome analysis fieldset . . . . .	313
14.5.9	Filter fieldset . . . . .	313
14.5.10	Understanding the output . . . . .	314
	14.5.10.1 Distance matrix . . . . .	314
	14.5.10.2 Unique strains . . . . .	314
14.6	BLAST . . . . .	314
	14.6.1 Include in results table fieldset . . . . .	317
	14.6.2 Parameters fieldset . . . . .	318
	14.6.3 No matches . . . . .	318
	14.6.4 Filter fieldset . . . . .	319
14.7	BURST . . . . .	319
14.8	Codon usage . . . . .	323
14.9	Unique combinations . . . . .	326
14.10	Polymorphisms . . . . .	328
14.11	Gene Presence . . . . .	330
	14.11.1 Pivot Table . . . . .	333
	14.11.2 Heatmap . . . . .	335
14.12	GrapeTree . . . . .	336
<b>15</b>	<b>Data export plugins</b>	<b>341</b>
15.1	Isolate record export . . . . .	341
	15.1.1 Advanced options . . . . .	343
	15.1.2 Molecular weight calculation . . . . .	343
15.2	Sequence export . . . . .	344
	15.2.1 Aligning sequences . . . . .	346
15.3	Profile export . . . . .	347
15.4	Contig export . . . . .	348
	15.4.1 Filtering by tagged status of contigs . . . . .	351
<b>16</b>	<b>Submitting data using the submission system</b>	<b>353</b>
16.1	Registering a user account . . . . .	354
16.2	Allele submission . . . . .	354
	16.2.1 Start . . . . .	354
	16.2.2 Select the submission locus . . . . .	355
	16.2.3 Enter details of sequencing method . . . . .	356
	16.2.4 Paste in sequence(s) . . . . .	357
	16.2.5 Add message to curator . . . . .	359
	16.2.6 Add supporting files . . . . .	361
	16.2.7 Finalize submission . . . . .	364
16.3	Profile submission . . . . .	365
	16.3.1 Start . . . . .	365
	16.3.2 Paste in profile(s) . . . . .	367
	16.3.3 Add message to curator . . . . .	368
	16.3.4 Add supporting files . . . . .	369
	16.3.5 Finalize submission . . . . .	370
16.4	Isolate submission . . . . .	371
	16.4.1 Start . . . . .	371
	16.4.2 Paste in isolate data . . . . .	372
	16.4.3 Add message to curator . . . . .	374
	16.4.4 Add supporting files . . . . .	375
	16.4.5 Finalize submission . . . . .	377

16.5	Genome submission . . . . .	379
16.6	Removing submissions from your notification list . . . . .	379
<b>17</b>	<b>RESTful Application Programming Interface (API)</b>	<b>381</b>
17.1	Passing additional/optional parameters . . . . .	381
17.2	Paging using request headers . . . . .	381
17.3	Resources . . . . .	382
17.3.1	GET / or /db - List site resources . . . . .	383
17.3.2	GET /db/{database} - List database resources . . . . .	384
17.3.3	GET /db/{database}/classification_schemes - List classification schemes . . . . .	384
17.3.4	GET /db/{database}/classification_schemes/{classification_scheme_id} - Retrieve classification scheme information and groups . . . . .	384
17.3.5	GET /db/{database}/classification_schemes/{classification_scheme_id}/groups/{group_id} - List isolates belonging to a classification scheme group . . . . .	385
17.3.6	GET /db/{database}/loci - List loci . . . . .	385
17.3.7	GET /db/{database}/loci/{locus} - Retrieve locus record . . . . .	386
17.3.8	GET /db/{database}/loci/{locus}/alleles - Retrieve list of alleles defined for a locus . . . . .	387
17.3.9	GET /db/{database}/loci/{locus}/alleles_fasta - Download alleles in FASTA format . . . . .	388
17.3.10	GET /db/{database}/loci/{locus}/alleles/{allele_id} - Retrieve full allele information . . . . .	388
17.3.11	POST /db/{database}/loci/{locus}/sequence - Query sequence to identify allele . . . . .	389
17.3.12	POST /db/{database}/sequence - Query sequence to identify allele without specifying locus . . . . .	390
17.3.13	GET /db/{database}/sequences - Get summary of defined sequences . . . . .	390
17.3.14	GET /db/{database}/schemes - List schemes . . . . .	391
17.3.15	GET /db/{database}/schemes/{scheme_id} - Retrieve scheme information . . . . .	391
17.3.16	GET /db/{database}/schemes/{scheme_id}/loci - Retrieve scheme loci . . . . .	392
17.3.17	GET /db/{database}/schemes/{scheme_id}/fields/{field} - Retrieve information about scheme field . . . . .	392
17.3.18	GET /db/{database}/schemes/{scheme_id}/profiles - List allelic profiles defined for scheme . . . . .	392
17.3.19	GET /db/{database}/schemes/{scheme_id}/profiles_csv - Download allelic profiles in CSV (tab-delimited) format . . . . .	393
17.3.20	GET /db/{database}/schemes/{scheme_id}/profiles/{profile_id} - Retrieve allelic profile record . . . . .	394
17.3.21	POST /db/{database}/schemes/{scheme_id}/sequence - Query sequence to extract allele designations/fields for a scheme . . . . .	394
17.3.22	GET /db/{database}/isolates - Retrieve list of isolate records . . . . .	395
17.3.23	GET /db/{database}/isolates/{isolate_id} - Retrieve isolate record . . . . .	396
17.3.24	GET /db/{database}/isolates/{isolate_id}/allele_designations - Retrieve list of allele designation records . . . . .	397
17.3.25	GET /db/{database}/isolates/{isolate_id}/allele_designations/{locus} - Retrieve full allele designation record . . . . .	397
17.3.26	GET /db/{database}/isolates/{isolate_id}/allele_ids - Retrieve allele identifiers . . . . .	398
17.3.27	GET /db/{database}/isolates/{isolate_id}/schemes/{scheme_id}/allele_designations - Retrieve scheme allele designation records . . . . .	399
17.3.28	GET /db/{database}/isolates/{isolate_id}/schemes/{scheme_id}/allele_ids - Retrieve list of scheme allele identifiers . . . . .	399
17.3.29	GET /db/{database}/isolates/{isolate_id}/contigs - Retrieve list of contigs . . . . .	399
17.3.30	GET /db/{database}/isolates/{isolate_id}/contigs_fasta - Download contigs in FASTA format . . . . .	400
17.3.31	GET /db/{database}/isolates/{isolate_id}/history - Retrieve isolate update history . . . . .	400
17.3.32	POST /db/{database}/isolates/search - Search isolate database . . . . .	401
17.3.33	GET /db/{database}/contigs/{contig_id} - Retrieve contig record . . . . .	401
17.3.34	GET /db/{database}/fields - Retrieve list of isolate provenance field descriptions . . . . .	402
17.3.35	GET /db/{database}/fields/{field} - Retrieve values set for a provenance field . . . . .	403
17.3.36	GET /db/{database}/users/{user_id} - Retrieve user information . . . . .	403
17.3.37	GET /db/{database}/projects - Retrieve list of projects . . . . .	404

17.3.38	GET /db/{database}/projects/{project_id} - Retrieve project information . . . . .	404
17.3.39	GET /db/{database}/projects/{project_id}/isolates - Retrieve list of isolates belonging to a project . . . . .	404
17.3.40	GET /db/{database}/submissions - retrieve list of submissions . . . . .	405
17.3.41	POST /db/{database}/submissions - create new submission . . . . .	405
17.3.42	GET /db/{database}/submissions/{submission_id} - Retrieve submission record . . . . .	406
17.3.43	DELETE /db/{database}/submissions/{submission_id} - Delete submission record . . . . .	408
17.3.44	GET /db/{database}/submissions/{submission_id}/messages - Retrieve submission correspondence . . . . .	408
17.3.45	POST /db/{database}/submissions/{submission_id}/messages - Add submission correspondence . . . . .	408
17.3.46	GET /db/{database}/submissions/{submission_id}/files - Retrieve list of supporting files uploaded for submission . . . . .	409
17.3.47	POST /db/{database}/submissions/{submission_id}/files - Upload submission supporting file . . . . .	409
17.3.48	GET /db/{database}/submissions/{submission_id}/files/{filename} - Download submission supporting file . . . . .	409
17.3.49	DELETE /db/{database}/submissions/{submission_id}/files/{filename} - Delete submission supporting file . . . . .	409
17.4	Authentication . . . . .	410
17.4.1	Developer sign up to get a consumer key . . . . .	410
17.4.2	Getting a request token . . . . .	410
17.4.3	Getting user authorization . . . . .	411
17.4.4	Getting an access token . . . . .	412
17.4.5	Getting a session token . . . . .	413
17.4.6	Accessing protected resources . . . . .	413
<b>18</b>	<b>Frequently asked questions (FAQs)</b> . . . . .	<b>415</b>
18.1	General . . . . .	415
18.2	Installation . . . . .	416
18.3	Administration . . . . .	416
<b>19</b>	<b>Appendix</b> . . . . .	<b>419</b>
19.1	Query operators . . . . .	419
19.2	Sequence tag flags . . . . .	420
19.3	Allele sequence flags . . . . .	421
<b>20</b>	<b>Database schema</b> . . . . .	<b>423</b>

*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/>.





### 1.1 BIGSdb

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

### 1.2 Loci

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

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

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

### 1.3 Alleles

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

When a specific allele of a locus is identified within the sequence data of an isolate record, the allele designation, i.e. identifier, is associated with the isolate record. This efficiently stores the sequence variation found within an isolate.

Two isolates with the same allele designation for a locus have identical sequences at that locus. Once the sequence variation within a genome has been reduced to a series of allele designations, genomes can be efficiently compared by identifying which loci vary between them.

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

## 1.4 Schemes

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

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

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

## 1.5 Profiles

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

## 1.6 Classification groups

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

## 1.7 Sequence tags

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

- sequence bin id - this identifies a particular contig
- locus name
- start position
- end position

- flag to indicate if sequence is reversed
- flag to indicate if sequence is complete and does not continue off the end of the contig

## 1.8 Sets

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

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.



### 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.

- [Archive::Zip](#) - Used to upload to iTOL.

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

### 2.1.3 Optional packages

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

- `ChartDirector` (Perl) - 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 javascript directory to the root directory of your website, i.e. accessible from [http://your\\_website/javascript/](http://your_website/javascript/).
5. Copy the css directory to root directory of your website, i.e. accessible from [http://your\\_website/css/](http://your_website/css/).
6. Copy the webfonts directory to the root directory of your website, i.e. accessible from [http://your\\_website/webfonts/](http://your_website/webfonts/).
7. Copy the images directory to the root directory of your website, i.e. accessible from [http://your\\_website/images/](http://your_website/images/).
8. Copy the contents of the conf directory to `/etc/bigsdb/`. Check the paths of helper applications and database names in the `bigsdb.conf` file and modify for your system.

9. Create a PostgreSQL database user called apache - this should not have any special priveleges. First you will need to log in as the postgres user:

```
sudo su postgres
```

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

```
createuser apache
```

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

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

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

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

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

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

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

## 3.2 Configuring PostgreSQL

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

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

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

### 3.2.1 pg\_hba.conf

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

# TYPE  DATABASE  USER  CIDR-ADDRESS  METHOD

# "local" is for Unix domain socket connections only
local    all             all                                ident map=mymap
```

(continues on next page)



(continued from previous page)

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

### 3.2.2 pg\_ident.conf

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

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

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

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

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

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

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

## 3.4 Site-specific configuration

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

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

The dropdown menu can be customized by modifying the `menu_header.html` file located in /etc/bigsdb. Any HTML in this file will be inserted at the top of the menu. The included file displays the BIGSdb logo hyperlinked to the root of the web site. This file can be edited as you wish, or alternatively `menu_header.html` can be placed in the root directory of the web site - this will be used in preference to the version in /etc/bigsdb.

## 3.5 Setting up the offline job manager

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

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

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

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

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

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

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

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

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

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

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

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

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

Add the following to /etc/crontab::

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

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

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

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

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

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

## 3.6 Setting up the submission system

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

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

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

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

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

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

## 3.7 Setting up a site-wide user database

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

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

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

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

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

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

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

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

You need to [add the users database details](#) to each client database that will use it.

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

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

```
site_user_dbs=PubMLST|pubmlst_bigsdb_users
```

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

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

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

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

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

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

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

- Allow admin to merge user accounts.
- modify\_users
  - Allow admin to edit user details.

e.g.

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

#### 4. Specific *permissions can be set for curators* in individual databases:

- import\_site\_users
    - This allows the curator to import site users in to the database.
  - modify\_site\_users
    - You may not wish to do this! - It allows the curator of any database with this permission to change the details of a user that may be used on other databases on the site.
5. HTML header files can be defined for use when bigsdb.pl or bigscurate.pl are called withouth a database configuration, such as when a user is registering or modifying their user details. These files, site\_header.html, site\_footer.html, site\_curate\_header.html and site\_curate\_footer.html should be placed in the root directory of the web site.

## 3.8 Periodically delete temporary files

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

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

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

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

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

## 3.9 Prevent preference database getting too large

The preferences database stores user preferences for BIGSdb databases running on the site. Every user will have a globally unique identifier (guid) stored in this database along with a 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'" bigsdb_prefs
```

## 3.10 Log file rotation

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

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

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

## 3.11 Upgrading BIGSdb

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

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

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

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

## 3.12 Running the BIGSdb RESTful interface

BIGSdb has an Application Programming Interface (API) that allows third-party applications to access the data within the databases. The script that runs this is called bigsrest.pl. This is a Dancer2 application that can be run using a wide

range of options, e.g. as a stand-alone script, using Perl 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.

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

```
description "Start BIGSdb REST interface"
version "1.0"
author "Keith Jolley"

start on runlevel [12345]

## tell upstart we're creating a daemon
expect fork

script

exec su -s /bin/sh -c 'exec "$0" "$@"' bigsdb -- /usr/local/bin/plackup -a /var/rest/
↪bigsrest.pl -s Starman -E deployment

end script
```

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

```
sudo service bigsdb-rest start
```

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

```
[Unit]
Description=BIGSdb REST interface
After=network.target

[Service]
User=bigsdb
ExecStart=/usr/bin/plackup -a /var/rest/bigsrest.pl -s Starman -E deployment
Restart=always
```

(continues on next page)



(continued from previous page)

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

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

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

It can also be manually started by calling:

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

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

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

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

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

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

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

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



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](http://pubmlst.org/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_neisseria_isolates) is the URL of the front page of the PubMLST Neisseria isolate database whose configuration settings are stored in /etc/bigsdb/dbases/pubmlst\_neisseria\_isolates. This database sub-directory contains a number of files (hyperlinks lead to the files used on the Neisseria database):

- config.xml - the database configuration file. Fields defined here correspond to fields in the isolate table of the database.
- banner.html - optional file containing text that will appear as a banner within the database index pages. HTML markup can be used within this text.
- header.html - HTML markup that is inserted at the top of all pages. This can be used to set up site-specific menubars and logos.
- footer.html - HTML markup that is inserted at the bottom of all pages.
- curate\_header.html - HTML markup that is inserted at the top of all curator's interface pages.
- curate\_footer.html - HTML markup that is inserted at the bottom of all curator's interface pages.
- profile\_submit.html - HTML markup for text that is inserted in to the submission interface prior to profile submission finalization. This can be used to add specific instructions such as the requirement to make an isolate submission.
- allele\_submit.html - HTML markup for text that is inserted in to the submission interface prior to allele submission finalization. This can be used to add specific instructions such as the requirement to attach Sanger trace files.

The header and footer files can alternatively be placed in the root directory of the web site for site-wide use.

There are four additional files, site\_header.html, site\_footer.html, curate\_site\_header.html and curate\_site\_footer.html which are used when either bigsdb.pl or bigscurate.pl are called without a database configuration. These should be placed in the root directory of the web site.

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

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

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

## 4.3 XML configuration attributes used in config.xml

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

### 4.3.1 Isolate database XML attributes

Please note that database structure described by the field 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>
```

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

- **authentication**
  - Method of authentication: either ‘builtin’ or ‘apache’. See *user authentication*.
- **db**
  - Name of database on system.
- **dbtype**
  - Type of database: either ‘isolates’ or ‘sequences’.
- **description**
  - Description of database used throughout interface.
- **align\_limit**
  - Overrides the sequence export record alignment limit in the Sequence Export plugin. Default: ‘200’.
- **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.
- **BLAST**
  - Enable Blast plugin: either ‘yes’ or ‘no’. If no value is set then the plugin will not be available unless the all\_plugins attribute is set to ‘yes’. If the all\_plugins attribute is set to ‘yes’, the Blast plugin can be disabled by setting this attribute to ‘no’.
- **BURST**

- Enable BURST plugin: either ‘yes’ or ‘no’. If no value is set then the plugin will not be available unless the `all_plugins` attribute is set to ‘yes’. If the `all_plugins` attribute is set to ‘yes’, the BURST plugin can be disabled by setting this attribute to ‘no’.
- `cache_schemes`
  - Enable automatic refreshing of scheme field caches when batch adding new isolates: either ‘yes’ or ‘no’, default ‘no’.
  - See [scheme caching](#).
- `CodonUsage`
  - Enable Codon Usage plugin: either ‘yes’ or ‘no’. If no value is set then the plugin will not be available unless the `all_plugins` attribute is set to ‘yes’. If the `all_plugins` attribute is set to ‘yes’, the Codon Usage plugin can be disabled by setting this attribute to ‘no’.
- `codon_usage_limit`
  - Overrides the record limit for the Codon Usage plugin. Default: ‘500’.
- `contig_analysis_limit`
  - Overrides the isolate number limit for the Contig Export plugin. Default: ‘1000’.
- `ContigExport`
  - Enable contig export plugin: either ‘yes’ or ‘no’. If no value is set then the plugin will not be available unless the `all_plugins` attribute is set to ‘yes’. If the `all_plugins` attribute is set to ‘yes’, the contig export plugin can be disabled by setting this attribute to ‘no’.
- `curate_config`
  - The database configuration that should be used for curation if different from the current configuration. This is used when the submission system is being used so that curation links in the ‘Manage submissions’ pages for curators load the correct database configuration.
- `curate_link`
  - URL to curator’s interface, which can be relative or absolute. This will be used to create a link in the public interface dropdown menu.
- `curate_path_includes`
  - Partial path of the `bigscurate.pl` script used to curate the database. See user authentication.
- `curate_script`
  - Relative web path to curation script. Default ‘bigscurate.pl’ (version 1.11+).
  - This is only needed if automated submissions are enabled. If `bigscurate.pl` is in a different directory from `bigsdbs.pl`, you need to include the whole web path, e.g. `/cgi-bin/private/bigsdbs/bigscurate.pl`.
- `curator_home`
  - URL of curator’s index page, which can be relative or absolute. This will be used to add a link in the dropdown menu.
- `curators_only`
  - Set to ‘yes’ to prevent ordinary authenticated users having access to database configuration. This is only effective if `read_access` is set to ‘authenticated\_users’. This may be useful if you have different configurations for curation and querying with some data hidden in the configuration used by standard users. Default ‘no’.
- `daily_pending_submissions`

- Overrides the daily limit on pending submissions that a user can submit via the web submission system. Default: '15'.
- `daily_rest_submissions_limit`
  - Overrides the limit on number of submissions that can be made to the database via the RESTful interface. This is useful to prevent flooding of the submission system by aberrant scripts. Default: '100'.
- `default_access`
  - The default access to the database configuration, either 'allow' or 'deny'. If 'allow', then specific users can be denied access by creating a file called 'users.deny' containing usernames (one per line) in the configuration directory. If 'deny' then specific users can be allowed by creating a file called 'users.allow' containing usernames (one per line) in the configuration directory. See [default access](#).
- `default_private_records`
  - The default number of private isolate records that a user can upload. The user account must have a status of either 'submitter', 'curator', or 'admin'. This value is used to set the `private_quota` field when creating a new user record (which can be overridden for individual users). Changing it will not affect the quotas of existing users. Default: '0'.
- `default_seqdef_config`
  - Isolate databases only: Name of the default seqdef database configuration used with this database. Used to automatically fill in details when adding new loci.
- `default_seqdef_dbase`
  - Isolate databases only: Name of the default seqdef database used with this database. Used to automatically fill in details when adding new loci.
- `default_seqdef_script`
  - Isolate databases only: URL of BIGSdb script running the seqdef database (default: '/cgi-bin/bigsdb/bigsdb.pl').
- `eav_fields`
  - Name to call sparsely-populated fields. Default: 'phenotypic fields'.
- `export_limit`
  - Overrides the default allowed number of data points (isolates x columns) to export. Default: '25000000'.
- `fast_scan`
  - Sets whether fast mode scanning is enabled via the web interface. This will scan all loci together, using exemplar sequences. In cases where multiple loci are being scanned this should be significantly faster than the standard locus-by-locus scan, but it will take longer for the first results to appear. [Allele exemplars](#) should be defined if you enable this option. Set to 'yes' to enable. Default: 'no'.
- `fieldgroup1 - fieldgroup10`
  - Allows multiple fields to be queried as a group. Value should be the name of the group followed by a colon (:) followed by a comma-separated list of fields to group, e.g. `identifiers:id,strain,other_name`.
- `GenomeComparator`
  - Enable Genome Comparator plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the `all_plugins` attribute is set to 'yes'. If the `all_plugins` attribute is set to 'yes', the Genome Comparator plugin can be disabled by setting this attribute to 'no'.
- `genome_comparator_limit`
  - Overrides the isolate number limit for the Genome Comparator plugin. Default: '1000'.

- `genome_comparator_max_ref_loci`
  - Overrides the limit on number of loci allowed in a reference genome. Default: '10000'.
- `genome_comparator_threads`
  - The number of threads to use for data gathering (BLAST, database queries) to populate data structure for Genome Comparator analysis. You should not set this to less than 2 as this will prevent job cancelling due to the way isolates are queued. Default: '2'.
- `hide_unused_schemes`
  - Sets whether a scheme is shown in a main results table if none of the isolates on that page have any data for the specific scheme: either 'yes' or 'no', default 'no'.
- `host`
  - Host name/IP address of machine hosting isolate database, default 'localhost'.
- `job_priority`
  - Integer with default job priority for offline jobs (default:5).
- `job_quota`
  - Integer with number of offline jobs that can be queued or currently running for this database.
- `labelfield`
  - Field that is used to describe record in isolate info page, default 'isolate'.
- `locus_aliases`
  - Display locus aliases and use them in dropdown lists by default: must be either 'yes' or 'no', default 'no'. This option can be overridden by a user preference.
- `locus_superscript_prefix`
  - Superscript the first letter of a locus name if it is immediately following by an underscore, e.g. `f_abcZ` would be displayed as `fabcZ` within the interface: must be either 'yes' or 'no', default 'no'. This can be used to designate gene fragments (or any other meaning you like).
- `maindisplay_aliases`
  - Default setting for whether isolates aliases are displayed in main results tables: either 'yes' or 'no', default 'no'. This setting can be overridden by individual user preferences.
- `Microreact`
  - Enable Microreact plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the `all_plugins` attribute is set to 'yes'. If the `all_plugins` attribute is set to 'yes', the Microreact plugin can be disabled by setting this attribute to 'no'. Note that for the plugin to be active, a country field containing a defined list of allowed values and an integer year field must be defined in the isolates table.
- `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'.
- PhyloViz
  - Enable third party PhyloViz plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all\_plugins attribute is set to 'yes'. If the all\_plugins attribute is set to 'yes', the PhyloViz plugin can be disabled by setting this attribute to 'no'.
- port
  - Port number that the isolate host is listening on, default '5432'.
- privacy
  - Displays E-mail address for sender in isolate information page if set to 'no'. Default 'yes'.
- public\_login
  - Optionally allow users to log in to a public database - this is useful as any jobs will be associated with the user and their preferences will also be linked to the account. Set to 'no' to disable. Default 'yes'.
- query\_script
  - Relative web path to bigsdb script. Default 'bigsdb.pl' (version 1.11+).
  - This is only needed if automated submissions are enabled. If bigsdb.pl is in a different directory from bigscurate.pl, you need to include the whole web path, e.g. /cgi-bin/bigsdb/bigsdb.pl.
- read\_access
  - Describes who can view data: either 'public' for everybody or 'authenticated\_users' for anybody who has been able to log in. Default 'public'.
- recommended\_schemes
  - Comma-separated list of recommended schemes to suggest to Genome Comparator users. If lots of schemes are defined, a user may be tempted to click 'All loci' and this may not be the best option. Populating this attribute, results in an additional list of preferred schemes that can be chosen.
- related\_databases
  - Semi-colon separated list of links to related BIGSdb databases on the system. This should be in the form of database configuration name followed by a 'l' and the description, e.g. 'pubmlst\_neisseria\_seqdeflSequence and profile definitions'. This is used to populate the dropdown menu.
- remote\_contigs
  - Optionally allow the use of remote contigs. These are stored in a remote BIGSdb database, accessible via the RESTful API. Set to 'yes' to enable.
- rMLSTSpecies
  - Enable rMLST Species identifier plugin: either 'yes' or 'no'. If no value is set then the plugin will not be available unless the all\_plugins attribute is set to 'yes'. If the all\_plugins attribute is set to 'yes', the plugin can be disabled by setting this attribute to 'no'. Note that for the plugin to be active, a country field containing a defined list of allowed values and an integer year field must be defined in the isolates table.
- script\_path\_includes
  - Partial path of the bigsdb.pl script used to access the database. See [user authentication](#).
- SeqbinBreakdown

- Enable Sequence bin breakdown plugin: either ‘yes’ or ‘no’. If no value is set then the plugin will not be available unless the `all_plugins` attribute is set to ‘yes’. If the `all_plugins` attribute is set to ‘yes’, the plugin can be disabled by setting this attribute to ‘no’. Note that for the plugin to be active, a country field containing a defined list of allowed values and an integer year field must be defined in the isolates table.
- `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 `bigsdbs.conf` file (for site-wide configuration) or within the system attribute of `config.xml`.
- `submissions_deleted_days`
  - Overrides the default number of days before closed submissions are deleted from the system. Default: ‘90’.
- `TagStatus`
  - Enable Tag status plugin: either ‘yes’ or ‘no’. If no value is set then the plugin will not be available unless the `all_plugins` attribute is set to ‘yes’. If the `all_plugins` attribute is set to ‘yes’, the plugin can be disabled by setting this attribute to ‘no’. Note that for the plugin to be active, a country field containing a defined list of allowed values and an integer year field must be defined in the isolates table.
- `tblastx_tagging`
  - Sets whether tagging can be performed using TBLASTX: either ‘yes’ or ‘no’, default ‘no’.
- `total_pending_submissions`
  - Overrides the total limit on pending submissions that a user can submit via the web submission system. Default: ‘20’.
- `user`
  - Username for access to isolates database, default ‘apache’.
- `user_job_quota`
  - Integer with number of offline jobs that can be queued or currently running for this database by any specific user - this parameter is only effective if users have to log in.

- `user_projects`
  - Sets whether authenticated users can create their own projects in order to group isolates: either 'yes' or 'no', default 'no'.
- `view`
  - Database view containing isolate data, default 'isolates'.
- `views`
  - Comma-separated list of views of the isolate table defined in the database. This is used to set a view for a set, or to restrict loci or schemes to a subset of isolate data.
- `webroot`
  - URL of web root, which can be relative or absolute. This is used to provide a hyperlinked item in the dropdown menu. Default '/'.

```
<field>
```

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

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

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

- Sets if field is displayed in the main table after a database search, ‘yes’ or ‘no’, default ‘yes’. This setting can be overridden by individual user preferences.
- max
  - Maximum value for integer and date types. Special values such as CURRENT\_YEAR and CURRENT\_DATE can be used.
- min
  - Minimum value for integer and date types.
- no\_curate
  - Setting this will hide the field in the curator interface and prevent it from being manually modified. This is useful for fields that are populated by automated scripts or database triggers. Can be ‘yes’ or ‘no’, default ‘no’.
- 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.

#### 4.3.1.1 Special values

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

- CURRENT\_DATE: current date in yyyy-mm-dd format
- CURRENT\_YEAR: the 4 digit value of the current year

`<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>
```

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

- **authentication**
  - Method of authentication: either ‘builtin’ or ‘apache’. See *user authentication*.
- **db**
  - Name of database on system.
- **dbtype**
  - Type of database: either ‘isolates’ or ‘sequences’.
- **description**
  - Description of database used throughout interface.
- **align\_limit**
  - Overrides the sequence export record alignment limit in the Sequence Export plugin. Default: ‘200’.
- **allele\_comments**
  - Enable comments on allele sequences: either ‘yes’ or ‘no’, default ‘no’.
  - This is not enabled by default to discourage the practice of adding isolate information to allele definitions (this sort of information belongs in an isolate database).
- **allele\_flags**
  - Enable flags to be set for alleles: either ‘yes’ or ‘no’, default ‘no’.
- **BURST**
  - Enable BURST plugin: either ‘yes’ or ‘no’. If no value is set then the plugin will not be available unless the all\_plugins attribute is set to ‘yes’. If the all\_plugins attribute is set to ‘yes’, the BURST plugin can be disabled by setting this attribute to ‘no’.
- **curate\_config**
  - The database configuration that should be used for curation if different from the current configuration. This is used when the submission system is being used so that curation links in the ‘Manage submissions’ pages for curators load the correct database configuration.
- **curate\_path\_includes**
  - Partial path of the bigscurate.pl script used to curate the database. See *user authentication*.
- **curate\_script**
  - Relative web path to curation script. Default ‘bigscurate.pl’ (version 1.11+).
  - This is only needed if automated submissions are enabled. If bigscurate.pl is in a different directory from bigsdb.pl, you need to include the whole web path, e.g. /cgi-bin/private/bigsdb/bigscurate.pl.
- **curator\_home**
  - URL of curator’s index page, which can be relative or absolute. This will be used to add a link in the dropdown menu.

- `curators_only`
  - Set to 'yes' to prevent ordinary authenticated users having access to database configuration. This is only effective if `read_access` is set to 'authenticated\_users'. This may be useful if you have different configurations for curation and querying with some data hidden in the configuration used by standard users. Default 'no'.
- `daily_pending_submissions`
  - Overrides the daily limit on pending submissions that a user can submit via the web submission system. Default: '15'.
- `daily_rest_submissions_limit`
  - Overrides the limit on number of submissions that can be made to the database via the RESTful interface. This is useful to prevent flooding of the submission system by aberrant scripts. Default: '100'.
- `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'.
- `exemplars`
  - Use exemplar sequences in the BLAST caches used for the sequence query pages. This is useful on larger databases as it speeds up the query significantly. *Exemplar alleles* *MUST* be defined otherwise sequence queries will fail. 'yes' or 'no', default 'no'.
- `isolate_database`
  - The config name of the isolate database. This is used to provide a link to isolate submissions. You also need to set `isolate_submissions="yes"`.
- `isolate_submissions`
  - Set to yes to provide a link to isolate submissions. The `isolate_database` attribute also needs to be set. Default: 'no'.
- `job_priority`
  - Integer with default job priority for offline jobs (default:5).
- `job_quota`
  - Integer with number of offline jobs that can be queued or currently running for this database.
- `profile_submissions`
  - Enable profile submissions (automated submission system): either 'yes' or 'no', default 'no' (version 1.11+).
  - To enable, you will also need to set `submissions="yes"`. By default, profile submissions are disabled since generally new profiles should be accompanied by representative isolate data, and the profile can be extracted from that.
- `public_login`
  - Optionally allow users to log in to a public database - this is useful as any jobs will be associated with the user and their preferences will also be linked to the account. Set to 'no' to disable. Default 'yes'.
- `query_script`

- Relative web path to bigsdb script. Default 'bigsdb.pl' (version 1.11+).
- This is only needed if automated submissions are enabled. If bigsdb.pl is in a different directory from bigscurate.pl, you need to include the whole web path, e.g. /cgi-bin/bigsdb/bigsdb.pl.
- read\_access
  - Describes who can view data: either 'public' for everybody, or 'authenticated\_users' for anybody who has been able to log in. Default 'public'.
- related\_databases
  - Semi-colon separated list of links to related BIGSdb databases on the system. This should be in the form of database configuration name followed by a 'l' and the description, e.g. 'pubmlst\_neisseria\_isolates|isolates'. This is used to populate the dropdown menu.
- script\_path\_includes
  - Partial path of the bigsdb.pl script used to access the database. See [user authentication](#).
- seq\_export\_limit
  - Overrides the sequence export limit (records x loci) in the Sequence Export plugin. Default: '1000000'.
- sets
  - Use [sets](#): either 'yes' or 'no', default 'no'.
- set\_id
  - Force the use of a specific set when accessing database via this XML configuration: Value is the name of the set.
- submissions
  - Enable automated submission system: either 'yes' or 'no', default 'no' (version 1.11+).
  - The curate\_script and query\_script paths should also be set, either in the bigsdb.conf file (for site-wide configuration) or within the system attribute of config.xml.
- submissions\_deleted\_days
  - Overrides the default number of days before closed submissions are deleted from the system. Default: '90'.
- total\_pending\_submissions
  - Overrides the total limit on pending submissions that a user can submit via the web submission system. Default: '20'.
- user\_job\_quota
  - Integer with number of offline jobs that can be queued or currently running for this database by any specific user - this parameter is only effective if users have to log in.
- webroot
  - URL of web root, which can be relative or absolute. This is used to provide a hyperlinked item in the dropdown menu. Default '/'.

## 4.4 Over-riding global defaults set in bigsdb.conf

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

- query\_script
  - Relative web path to bigsdb script.
- curate\_script
  - Relative web path to curation script.
- prefs\_db
  - The name of the preferences database.
- auth\_db
  - The name of the authentication database.
- tmp\_dir
  - Path to the web-accessible temporary directory.
- secure\_tmp\_dir
  - Path to the web-inaccessible (secure) temporary directory.
- ref\_db
  - The name of the references database.

## 4.5 Over-riding values set in config.xml

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

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

## 4.6 User authentication

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

### 4.6.1 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'`.

## 4.6.2 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'`.

## 4.7 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](http://your_website/cgi-bin/private/bigscurate.pl?db=test_db) (or wherever you have located your `bigscurate.pl` script).

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

```
pubmlst_bigsdb_neisseria_isolates      refs
pubmlst_bigsdb_neisseria_seqdef       profile_refs, sequence_refs, locus_refs
```

The script can be called as follows:

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

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

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

## 4.9 Configuring access to remote contigs

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

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

```
remote_contigs = "yes"
```

### 4.9.1 Setting up authentication

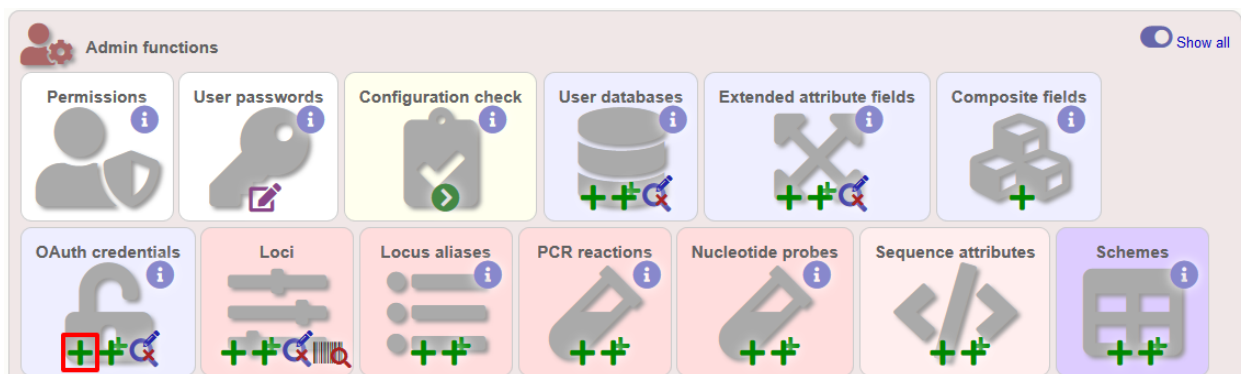
A client key for the BIGSdb remote contig manager needs to be generated. This can be done using the *create\_client\_credentials.pl* script, e.g.

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

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

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

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



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


PubMLST Database home Curator home Contents

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

### Add new OAuth credentials

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

Record	Action
base uri!: <input type="text" value="http://rest.pubmlst.org/db/pubmlst_rm1st_isolates"/> consumer key!: <input type="text" value="RylqmtjBbSAkyrZ5ZUyHzae"/> consumer secret!: <input type="text" value="U6B_^%HJP9N^1Ax2wemPPkmqrT_3%rX6pkgkfdho"/> access token!: <input type="text" value="eHTH8jMXsnyhIUvJZFmDF7nj5IZQJsd"/> access secret!: <input type="text" value="QODGPqKoSgVcGwdKBjufDsn4B1beX1"/> curator!: <input type="text" value="Keith Jolley (keith)"/> date entered!: <input type="text" value="2017-11-23"/> datestamp!: <input type="text" value="2017-11-23"/>	<input type="button" value="Reset"/> <input type="button" value="Submit"/>



## 4.9.2 Processing remote contigs

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

```
remote_contigs.pl --help
NAME
    process_remote_contigs.pl
    Download, check length and create checksum contigs stored as URIs

SYNOPSIS
    process_remote_contigs.pl --database NAME [options]

OPTIONS
--database NAME
    Database configuration name.

--exclude_isolates LIST
    Comma-separated list of isolate ids to ignore.

--exclude_projects LIST
    Comma-separated list of projects whose isolates will be excluded.

--help
    This help page.

--isolates LIST
    Comma-separated list of isolate ids to scan (ignored if -p used).

--isolate_list_file FILE
```

(continues on next page)

(continued from previous page)

```
File containing list of isolate ids (ignored if -i or -p used).

--min ID
    Minimum isolate id.

--max ID
    Maximum isolate id.

--projects LIST
    Comma-separated list of project isolates to scan.

--quiet
    Only display errors.
```

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

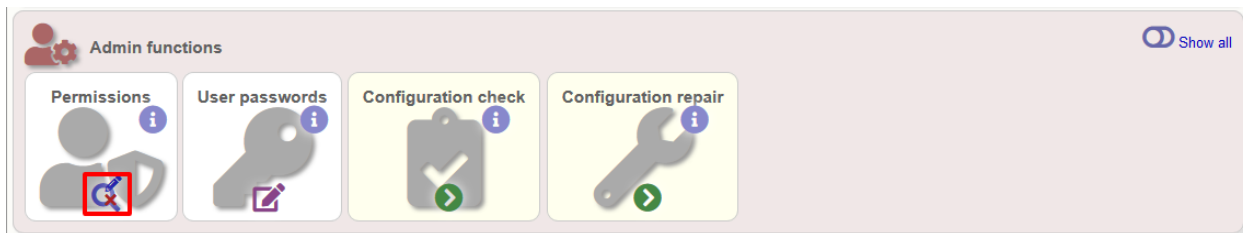
User groups allow submitter accounts to be grouped such that the submitter can edit isolates where the sender is either themselves or any member of a user group to which they belong.

## 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_locus_descriptions` - allowed to modify the description text and external hyperlinks used. Even with this setting, only loci for which a user is explicitly set as a curator can be modified.
- `modify_schemes` - allowed to add or modify schemes.
- `modify_sequences` - allowed to add sequences to the sequence bin (for isolate databases) or new allele definitions (for sequence definition databases).
- `modify_experiments` - define new experiments that can be used to group contigs uploaded to the sequence bin.
- `modify_isolates` - allowed to add or modify isolate records.
- `modify_projects` - allowed to create projects, modify their descriptions and add or remove isolate records to these.
- `modify_composites` - allowed to add or modify composite fields (fields made up of other fields, including scheme fields defined in external databases). Composite fields involve defining regular expressions that are evaluated by Perl - this can be dangerous so this permission should be granted with discretion.
- `modify_field_attributes` - allow user to create or modify secondary field attributes (lookup tables) for isolate record fields.
- `modify_value_attributes` - allow user to add or modify secondary field values for isolate record fields.
- `modify_probes` - allow user to define PCR or hybridization reactions to filter tag scanning.
- `tag_sequences` - allowed to tag sequences with locus information.
- `designate_alleles` - allowed to manually designate allele numbers for isolate records.
- `modify_profiles` - allowed to add or modify scheme profiles (only used in a sequence definitions database).
- `import_site_users` - allowed to import site users in to the database.
- `modify_site_users` - allowed to modify site user details (you may not want to this! The user account can be used by multiple databases on the site and any changes to user details will be seen throughout the site).

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



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

PubMLST Database home Curator home Contents

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

### Set curator permissions

Select curator(s)	Action
Clark, Stephen (SClark)	<input type="checkbox"/> Select
Clemence, Marianne (mclemence)	
Debech, Nadia (ndebech)	
Deghmane, Ala-Eddin (deghmane)	
Diallo, Kanny (kdiallo)	
Exley, Rachel (rexley)	
Feavers, Ian (ifeavers)	
Goodyer-Sait, Lily (lgoodyer)	

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

PubMLST Database home Curator home Contents

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

## Set curator permissions

Select curator(s) Action

Clark, Stephen (SClark)  
 Clemence, Marianne (mclemence)  
 Debech, Nadia (ndebech)  
 Deghmane, Ala-Eddin (degmane)  
 Diallo, Kanny (kdiallo)  
 Exley, Rachel (rexley)  
 Feavers, Ian (ifeavers)  
 Goodyer-Sait, Lily (lgoodyer)

All None

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. Attributes with a red background add restrictions.

Update permissions Action

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"/>
import site users	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
modify site users	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
only private	<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"/>

Update

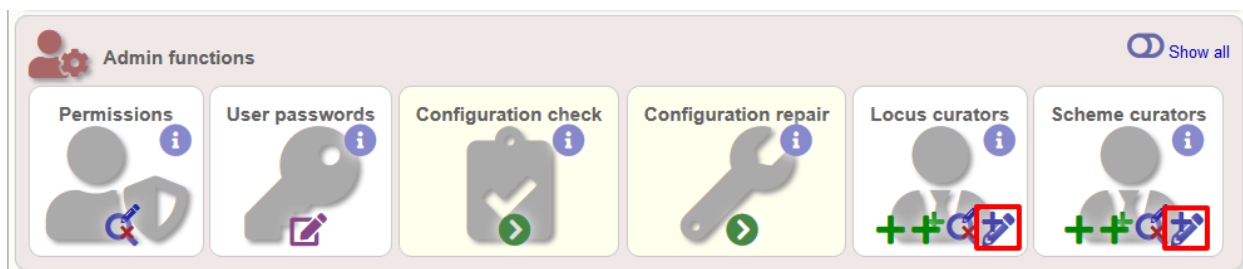
The 'disable access' option provides a quick way to disable access to a curator. This will not be selected by the 'All/None' buttons.

## 5.4 Locus and scheme permissions (sequence definition database)

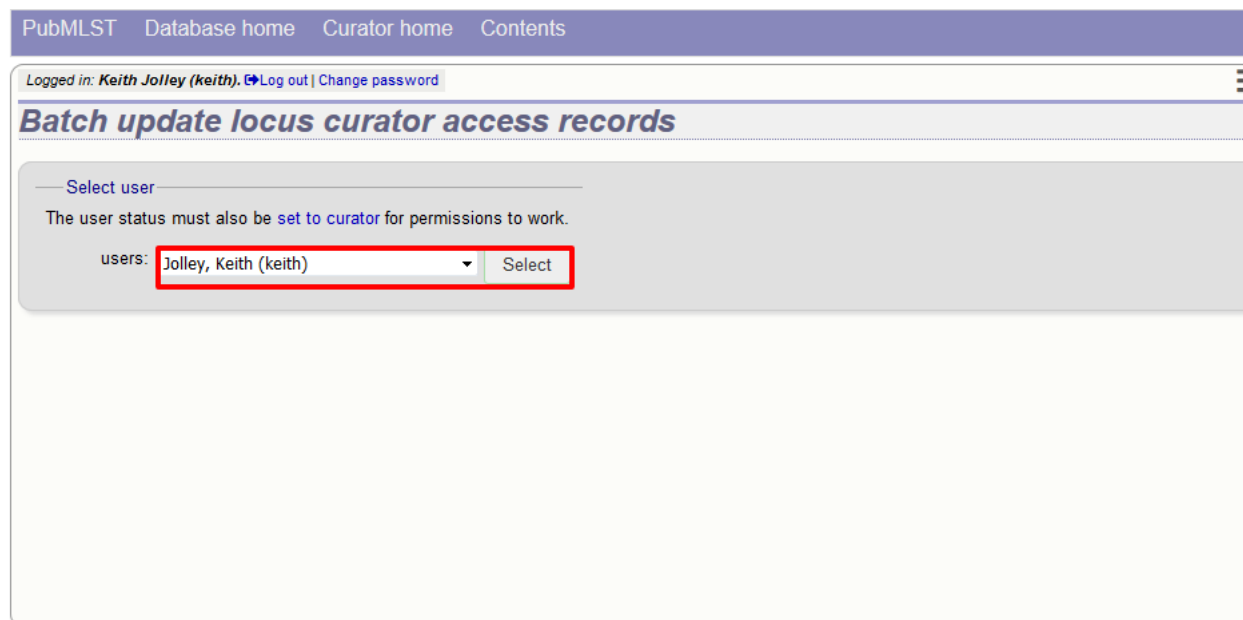
To be allowed to define alleles or scheme profiles, curators must be granted specific permission for each locus and scheme by adding their user id number to the 'locus curator' and 'scheme curator' lists.

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':





Select the curator from the list:



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

PubMLST Database home Curator home Contents

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
NEIS0001		abcZ
NEIS0004		adk
NEIS0005		aroE
NEIS0006		aspA
NEIS0007		carB
NEIS0008	→	dhpS
NEIS0009	←	fumC
NEIS0010		gdh
NEIS0011		glnA
NEIS0012		gpm
NEIS0013		mtgA
NEIS0014		NEIS1364

All None All None

☒ Hide curator name from public view

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

## 5.5 Controlling access

### 5.5.1 Restricting particular configurations to specific user accounts

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

In the config.xml, add the following directive:

```
default_access="deny"
```

This tells BIGSdb to deny access to anybody unless their account name appears within a file called users.allow within the config directory. The users.allow file should contain one username per line. You can also use a usergroups.allow file. This file should contain the names of user groups, the members of which are allowed access. The file should contain one user group name per line.

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

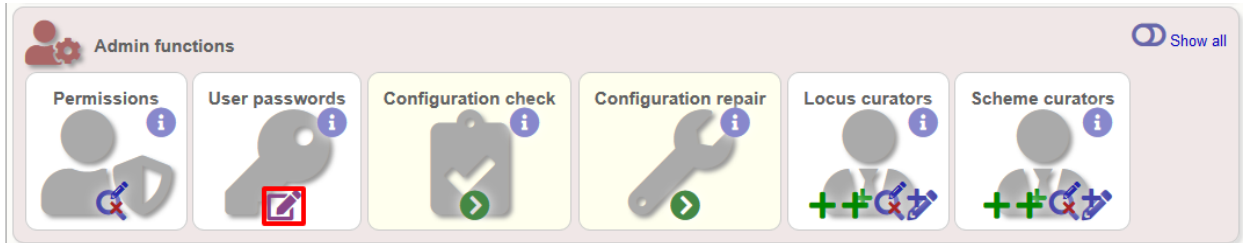
```
default_access="allow"
```

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

## 5.6 Setting user passwords

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

If you are an administrator or a curator with specific permission to change other users' passwords, you should see a 'User passwords' box in the admin section of the curator's index page. Click the 'Set passwords' link.



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

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. If the `system_flag` value is set to `'no'` then the plugin will be disabled even if the `all_plugins` attribute is set to `'yes'`.

## 5.9 Temporarily disabling database updates

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

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

### 5.9.1 Global

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

```
disable_updates=yes
```

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

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

## 5.9.2 Database-specific

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

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

## 5.10 Host mapping

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

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

<i>#Existing_host</i>	<i>Mapped_host</i>
server1	server2
localhost	server2

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

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

## 5.11 Improving performance

### 5.11.1 Use mod\_perl

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

### 5.11.2 Cache scheme definitions within an isolate database

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

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

```
update_scheme_caches.pl --database pubmlst_bigsdb_neisseria_isolates
```

This script creates indexed tables within the isolate database called temp\_scheme\_X and temp\_isolates\_scheme\_fields\_1 (where X is the scheme\_id). If these table aren't present, they are created as temporary tables every time a query is performed that requires a join against scheme definition data. This requires

importing all profile definitions from the definitions database and determining scheme field values for all isolates. This may sound like it would be slow but caching only has a noticeable effect once you have >5000 profiles.

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

```
update_scheme_caches.pl --help

NAME
    update_scheme_caches.pl - Update scheme field caches

SYNOPSIS
    update_scheme_caches.pl --database NAME [options]

OPTIONS

--database NAME
    Database configuration name.

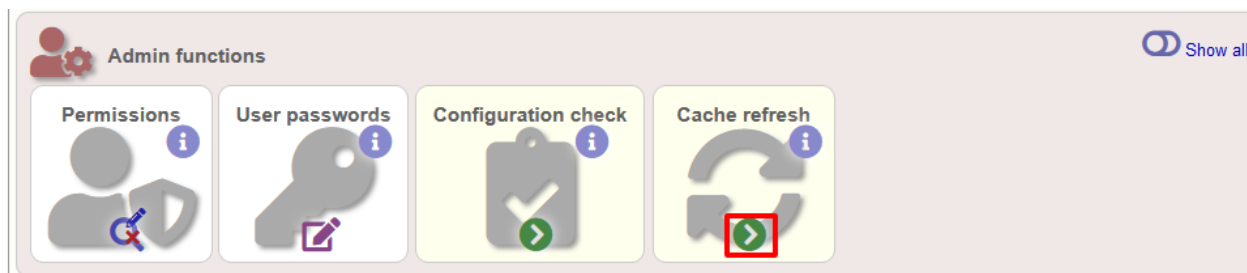
--help
    This help page.

--method METHOD
    Update method - the following values are allowed:
    full: Completely recreate caches
    incremental: Only add values for records not in cache.
    daily: Only add values for records not in cache updated today.
    daily_replace: Refresh values only for records updated today.

--quiet
    Don't output progress messages.

--schemes SCHEMES
    Comma-separated list of scheme ids to use.
    If left empty, all schemes will be updated.
```

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



You can also set `cache_schemes="yes"` in the system tag of `config.xml` to enable automatic refreshing of the caches (using the ‘daily’ method) when batch adding new isolates (you should still periodically run the `update_scheme_caches.pl` script via CRON to ensure any changes in the sequence definition database are picked up).

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

### 5.11.3 Use a ramdisk for the secure temporary directory

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

## 5.12 Dataset partitioning

### 5.12.1 Sets

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

**See also:**

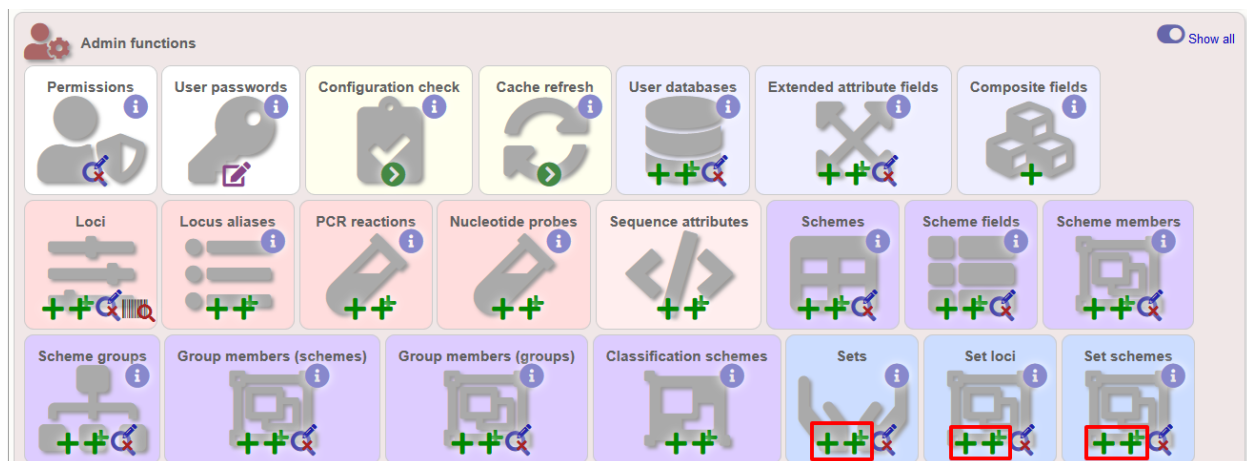
*Sets (concept)*

### 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:

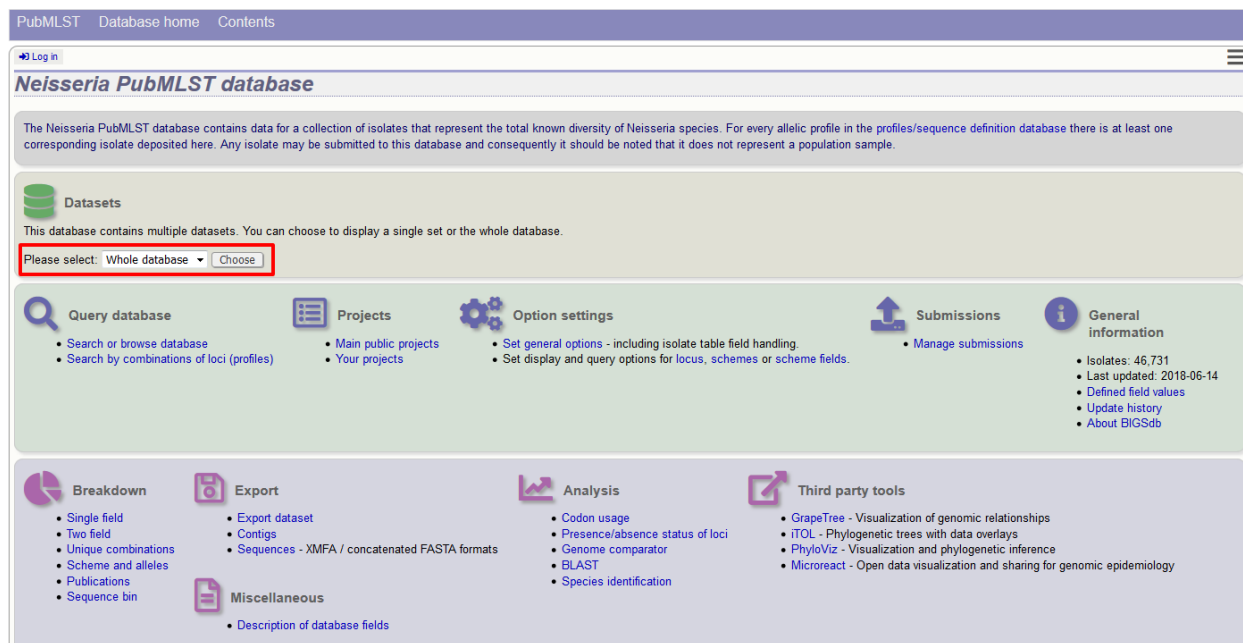
```
sets="yes"
```

With this attribute, the curation interface now has options to add sets, and then add loci or schemes to these sets. These functions are normally hidden, so you may need to click the ‘Show all’ toggle to display it.



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

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



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

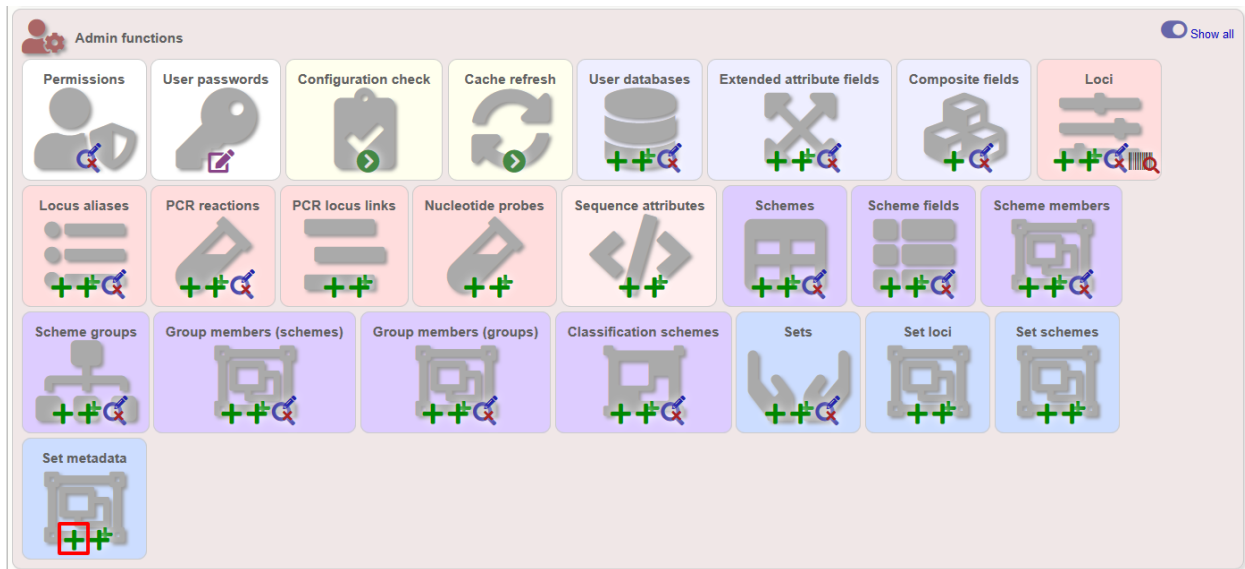
Additional metadata fields can be set within the XML configuration file. They are specified as belonging to a metaset by prefixing the field name with 'meta\_NAME:' where NAME is the name of the metaset, e.g.

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



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

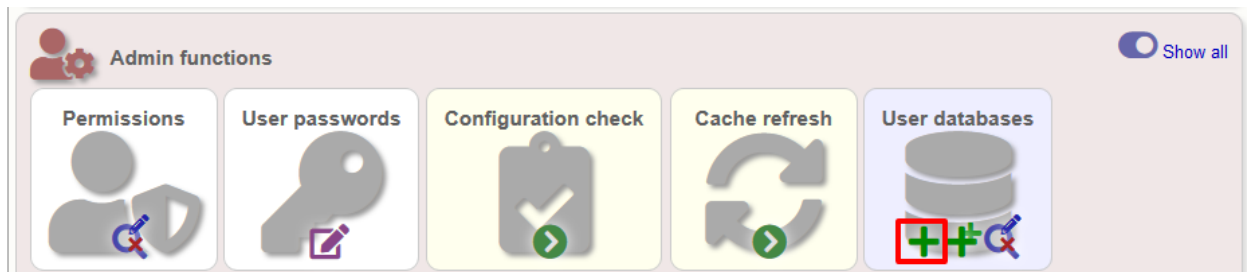
The `only_sets` attribute can be set to ‘yes’ to disable the option for ‘Whole database’ so that only sets can be viewed, e.g.

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

## 5.13 Setting a site-wide users database

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

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



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

PubMLST Database home Curator home Contents

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

## Add new user database

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

Record

id: 1

name: PubMLST Site/domain name

dbase name: pubmlst\_bigsdb\_users Name of the database holding user data

curator: Keith Jolley (keith)

datestamp: 2016-12-16

list order: [v]

auto registration: ☐ true ☐ false Allow user to register themselves for database

dbase host: [ ] IP address of database host

dbase port: [ ] Network port accepting database connections

dbase user: [ ]

dbase password: [ ]

Action

Reset Submit

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

## 5.14 Adding new loci

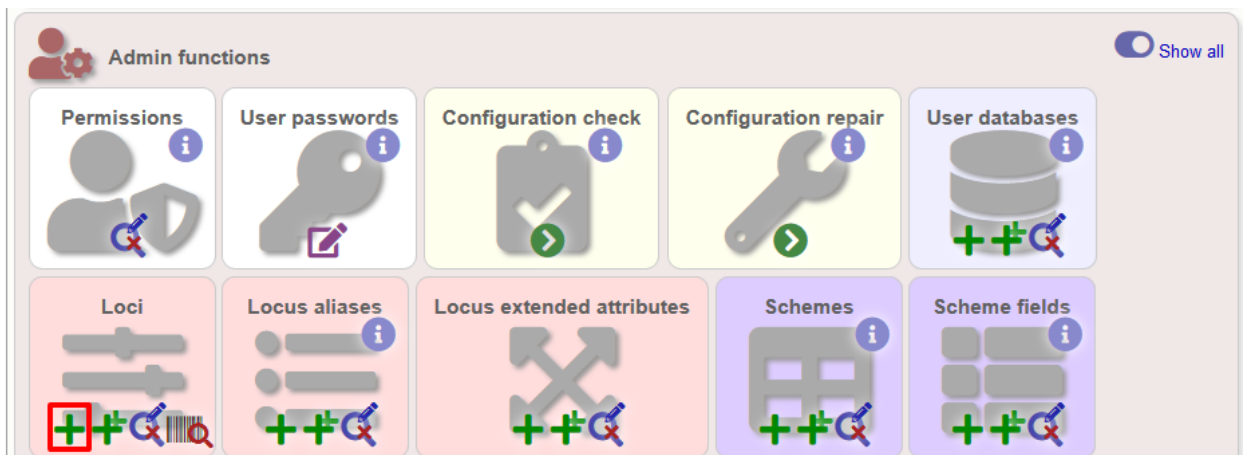
See also:

*Loci (concept)*

### 5.14.1 Sequence definition databases

#### 5.14.1.1 Single locus

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



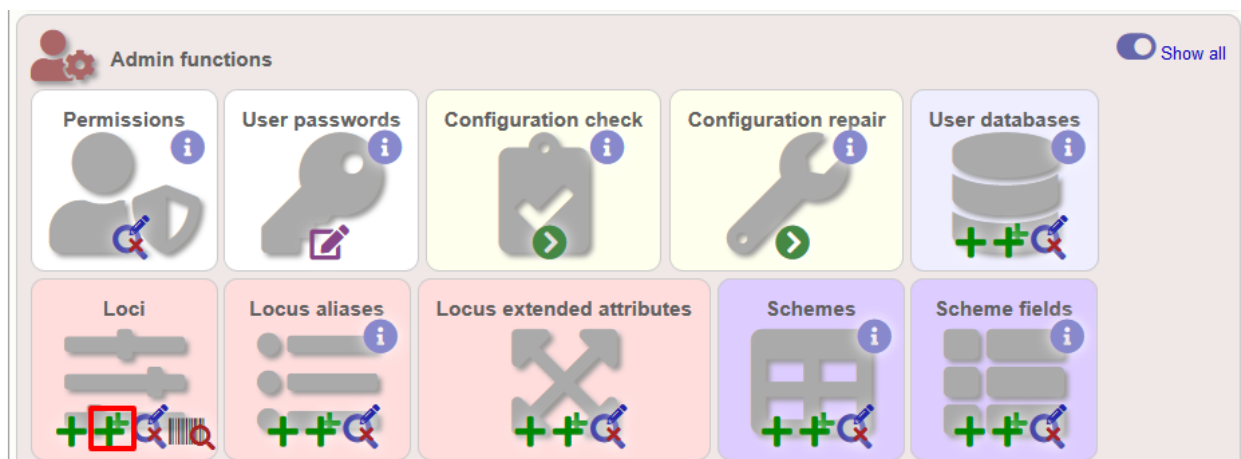
Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

- id - The name of the locus.
  - Allowed: any value starting with a letter, number or underscore.
- data\_type - Describes whether the locus is defined by nucleotide or peptide sequence.
  - Allowed: DNA/peptide.
- allele\_id\_format - The format for allele identifiers.
  - Allowed: integer/text.
- length\_varies - Sets whether alleles can vary in length.
  - Allowed: true/false.
- coding\_sequence - Sets whether the locus codes for a protein.
  - Allowed: true/false.
- formatted\_name - Name with HTML formatting (optional).
  - This allows you to add formatting such as bold, italic, underline and superscripting to locus names as they appear in the web interface.
  - Allowed: valid HTML.
- common\_name - The common name for the locus (optional).
  - Allowed: any value.
- formatted\_common\_name - Common name with HTML formatting (optional).
  - Allowed: valid HTML.
- allele\_id\_regex - **Regular expression** to enforce allele id naming (optional).
  - ^: the beginning of the string
  - \$: the end of the string
  - d: digit
  - D: non-digit
  - s: white space character
  - S: non white space character
  - w: alpha-numeric plus ‘\_’
  - .: any character
  - \*: 0 or more of previous character
  - +: 1 or more of previous character
  - e.g. ^Fd-d+\$ states that an allele name must begin with a F followed by a single digit, then a dash, then one or more digits, e.g. F1-12
- length - Standard length of locus (required if length\_varies is set to false).
  - Allowed: any integer.
- min\_length - Minimum length of locus (optional).
  - Allowed: any integer.
- max\_length - Maximum length of locus (optional).

- Allowed: any integer (larger than the minimum length).
- orf - Open reading frame of locus (optional).
  - 1-3 are the forward reading frame, 4-6 are the reverse reading frames.
  - Allowed: 1-6.
- genome\_position - The start position of the locus on a reference genome (optional).
  - Allowed: any integer.
- match\_longest - Specifies whether in a sequence query to only return the longest match (optional).
  - This is useful for some loci that can have some sequences shorter than others, e.g. peptide loci defining antigenic loops. This can lead to instances of one sequence being longer than another but otherwise being identical. In these cases, usually the longer sequence is the one that should be matched.
  - Allowed: true/false.
- full\_name - Full name of the locus (optional).
  - Allowed: any value.
- product - Name of gene product (optional).
  - Allowed: Any value.
- description - Description of the locus (optional).
  - Allowed: any value.
- aliases - Alternative names for the locus (optional).
  - Enter each alias on a separate line in the text box.
  - Allowed: any value.
- pubmed\_ids - PubMed ids of publications describing the locus (optional).
  - Enter each PubMed id on a separate line in the text box.
  - Allowed: any integer.
- links - Hyperlinks pointing to additional resources to display in the locus description (optional).
  - Enter each link on a separate line in the format with the URL first, followed by a | then the description (URL|description).

#### 5.14.1.2 Batch adding

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



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

PubMLST Database home Curator home Contents

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

### Batch insert loci

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

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

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

Action: Reset Submit

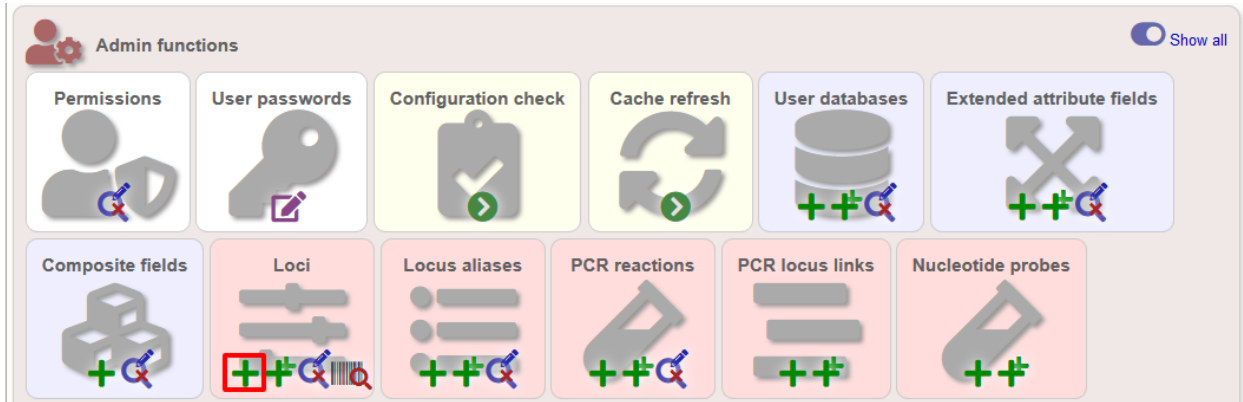
Fill in the spreadsheet using the fields described for *adding single loci*.

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

## 5.14.2 Isolate databases

### 5.14.2.1 Single locus

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



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

Logged in: Keith Jolley (keith) | Log out | Change password | Toggle: [icon] [icon] | Show tools

### Add new locus

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

**Record**

id: [text field]

data type: [dropdown: DNA]

allele id format: [dropdown: integer] !

length varies: [radio: true] [radio: false] !

coding sequence: [radio: true] [radio: false]

isolate display: [dropdown: allele only] !

main display: [radio: true] [radio: false] !

query field: [radio: true] [radio: false] !

analysis: [radio: true] [radio: false] !

curator: Keith Jolley (keith)

date entered: 2016-07-26

datestamp: 2016-07-26

formatted name: [text field] !

common name: [text field] !

formatted common name: [text field] !

allele id regex: [text field] !

length: [text field] !

orf: [dropdown] !

genome position: [text field] !

match longest: [radio: true] [radio: false] !

reference sequence: [text area]

pcr filter: [radio: true] [radio: false] ! Do NOT set to true unless you define PCR reactions linked to this locus.

probe filter: [radio: true] [radio: false] ! Do NOT set to true unless you define probe sequences linked to this locus.

dbase name: pubmist\_bigsdb\_neisseria\_seqdef ! Name of the database holding allele sequences

dbase host: [text field] ! IP address of database host

dbase port: [text field] ! Network port accepting database connections

dbase user: [text field] !

dbase password: [text field] !

dbase id: PUT\_LOCUS\_NAME\_HERE ! Name of locus in seqdef database

**Action**

Reset Submit

- id - The name of the locus

- Allowed: any value starting with a letter or underscore.
- data\_type - Describes whether the locus is defined by nucleotide or peptide sequence.
  - Allowed: DNA/peptide.
- allele\_id\_format - The format for allele identifiers.
  - Allowed: integer/text.
- length\_varies - Sets whether alleles can vary in length.
  - Allowed: true/false.
- coding\_sequence - Sets whether the locus codes for a protein.
  - Allowed: true/false.
- isolate\_display - Sets how alleles for this locus are displayed in a detailed isolate record - this can be overridden by user preference.
  - Allowed: allele only/sequence/hide.
- main\_display - Sets whether or not alleles for this locus are displayed in a main results table by default - this can be overridden by user preference.
  - Allowed: true/false.
- query\_field - Sets whether or not alleles for this locus can be used in queries by default - this can be overridden by user preference.
  - Allowed: true/false.
- analysis - Sets whether or not alleles for this locus can be used in analysis functions by default - this can be overridden by user preference.
  - Allowed: true/false.
- formatted\_name - Name with HTML formatting (optional).
  - This allows you to add formatting such as bold, italic, underline and superscripting to locus names as they appear in the web interface.
  - Allowed: valid HTML.
- common\_name - The common name for the locus (optional).
  - Allowed: any value.
- formatted\_common\_name - Common name with HTML formatting (optional).
  - Allowed: valid HTML.
- allele\_id\_regex - **Regular expression** to enforce allele id naming.
  - ^: the beginning of the string
  - \$: the end of the string
  - d: digit
  - D: non-digit
  - s: white space character
  - S: non white space character
  - w: alpha-numeric plus ‘\_’
  - .: any character



- \*: 0 or more of previous character
  - +: 1 or more of previous character
  - e.g. ^Fd-d+\$ states that an allele name must begin with a F followed by a single digit, then a dash, then one or more digits, e.g. F1-12
- length - Standard length of locus (required if length\_varies is set to false).
  - Allowed: any integer.
- orf - Open reading frame of locus (optional). 1-3 are the forward reading frame, 4-6 are the reverse reading frames.
  - Allowed: 1-6.
- genome\_position - The start position of the locus on a reference genome.
  - Allowed: any integer.
- match\_longest - Only select the longest exact match when tagging/querying.
  - This is useful for some loci that can have some sequences shorter than others, e.g. peptide loci defining antigenic loops. This can lead to instances of one sequence being longer than another but otherwise being identical. In these cases, usually the longer sequence is the one that should be matched.
  - Allowed: true/false.
- reference\_sequence - Sequence used by the automated sequence comparison algorithms to identify sequences matching this locus. **This is only used if a sequence definition database has not been set up for this locus.**
- pcr\_filter - Set to true if this locus is further defined by genome filtering using in silico PCR.
  - Allowed: true/false.
- probe\_filter - Set to true if this locus is further defined by genome filtering using in silico 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\_id - Name of locus in seqdef database. This is usually the same as the id field.
  - Allowed: any text (no spaces).

- `description_url` - The URL used to hyperlink to locus information in the isolate information page. This can either be a relative (e.g. `/cgi-bin/...`) or an absolute (containing `http://`) URL.
  - Allowed: any valid URL.
- `url` - The URL used to hyperlink to information about the allele. This can either be a relative or absolute URL. If `[?]` (including the square brackets) is included then this will be substituted for the allele value in the resultant URL. To link to the appropriate allele info page on a corresponding seqdef database you would need something like `/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_neisseria_seqdef&page=alleleInfo&locus=abcZ&allele_id=[?]`.
  - Allowed: any valid URL.
- `submission_template` - Sets whether or not a column for this locus is included in the Excel submission template.
  - Allowed: true/false (default: false)
- `view` - Restrict this scheme to only isolates contained in the specified database view. This option will only appear if the views attribute is set in the system tag. The view needs to have been defined in the database as a subset of the isolate table (usually filtered by the value of one or more of its fields).

## Using existing locus definition as a template

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

PubMLST Database home Curator home Contents

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

**Add new locus** [Show tools](#)

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

Record

id:!

data type:! DNA

allele id format:! integer

length varies:! ☐ true ☒ false

coding sequence:! ☒ true ☐ false

isolate display:! allele only

main display:! ☐ true ☒ false

query field:! ☒ true ☐ false

analysis:! ☒ true ☐ false

curator:! Keith Jolley (keith)

date entered:! 2018-06-18

datestamp:! 2018-06-18

formatted name:

common name:

formatted common name:

allele id regex:

length:

complete cds: ☐ true ☐ false

orf:

genome position:

match longest: ☐ true ☐ false

reference sequence:

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

PubMLST Database home Curator home Contents

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

## Add new locus

Hide tools

Please fill in the fields below - required fields are marked \*

**Record**

id: \*

data type: DNA

allele id format: integer

length varies: ☐ true ☒ false

coding sequence: ☒ true ☐ false

isolate display: allele only

main display: ☐ true ☒ false

query field: ☒ true ☐ false

analysis: ☒ true ☐ false

curator: Keith Jolley (keith)

date entered: 2018-06-18

date stamp: 2018-06-18

formatted name:

common name:

formatted common name:

allele id regex:

length:

complete cds: ☐ true ☒ false

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

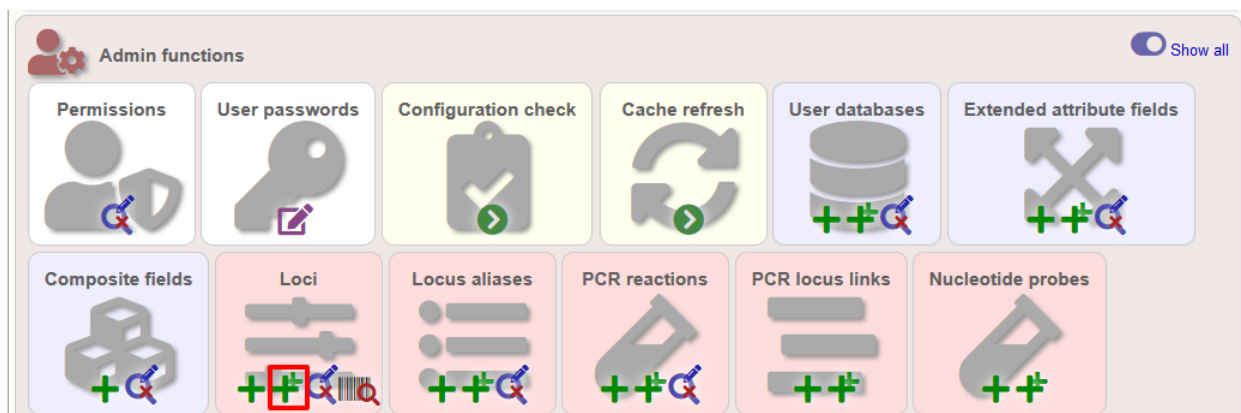
aliases:

Action

Complete the form and click 'Submit'.

### 5.14.2.2 Batch adding

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



Click the link to download an Excel template:

PubMLST Database home Curator home Contents

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.
- 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).

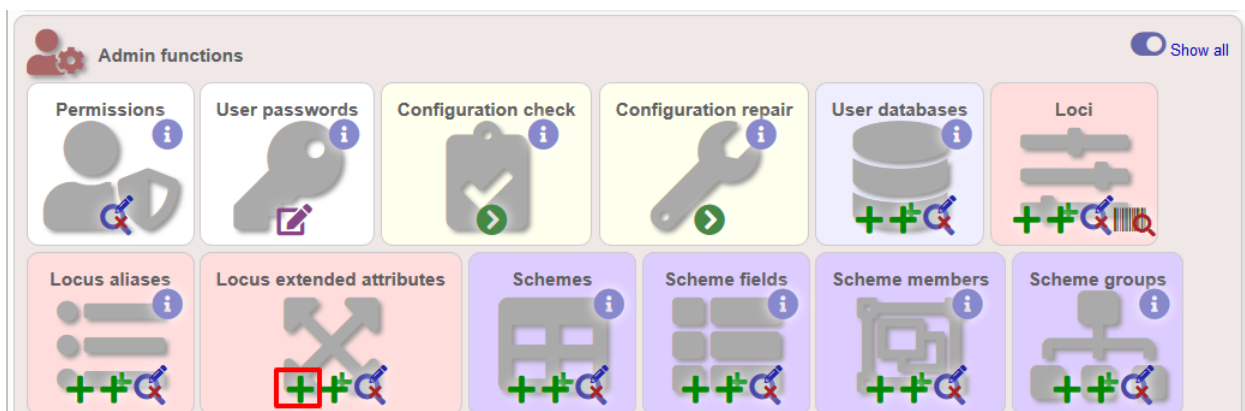
Action  
Reset Submit

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

## 5.15 Defining locus extended attributes

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

To add extended attributes for a locus, click add (+) locus extended attributes in the sequence definition database curator's interface contents page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



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

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

- +: 1 or more of previous character
- description - Description that will appear within the web form when adding new sequences (optional).
  - Allowed: any value.
- option\_list - Pipe (|) separated list of allowed values (optional).
- length - Maximum length of value (optional).
  - Allowed: any integer.
- field\_order - Integer that sets the position of the field within scheme values in any results (optional).
  - Allowed: any integer.

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

## 5.16 Defining schemes

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

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

To set up a new scheme, you need to:

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

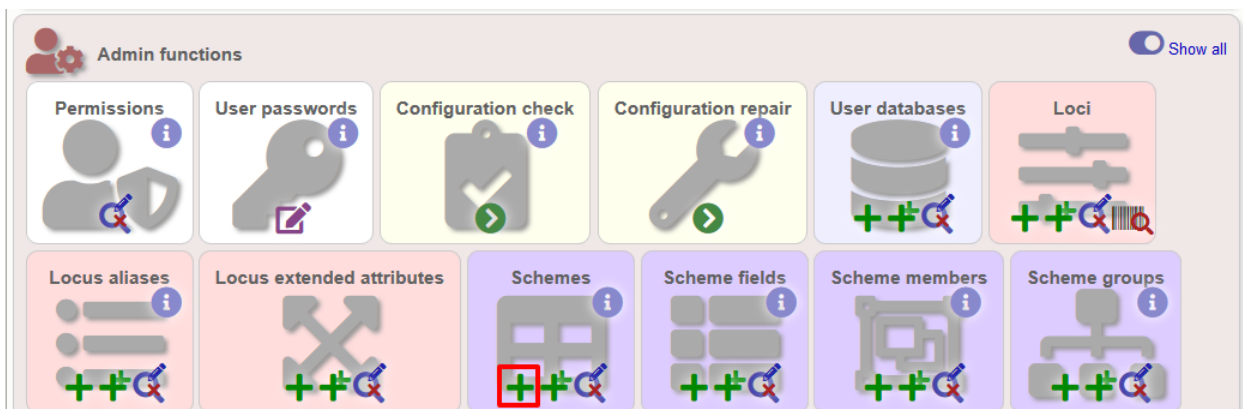
**See also:**

*Schemes (concept)*

### 5.16.1 Sequence definition databases

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



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



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

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

PubMLST Database home Curator home Contents


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:

name:  


curator: **Keith Jolley (keith)**

timestamp: **2018-06-18**

date entered: **2018-06-18**


description:

display order:

allow missing loci: ☐ true ☐ false  This is only relevant to schemes with primary key fields, e.g. MLST.

disable: ☐ true ☐ false Set to true to disable scheme. This can be overridden by user preference settings.

no submissions: ☐ true ☐ false Set to true to prevent submission of profiles of this scheme via the automated submission system.

flags:   Use CTRL/SHIFT click to select or deselect values

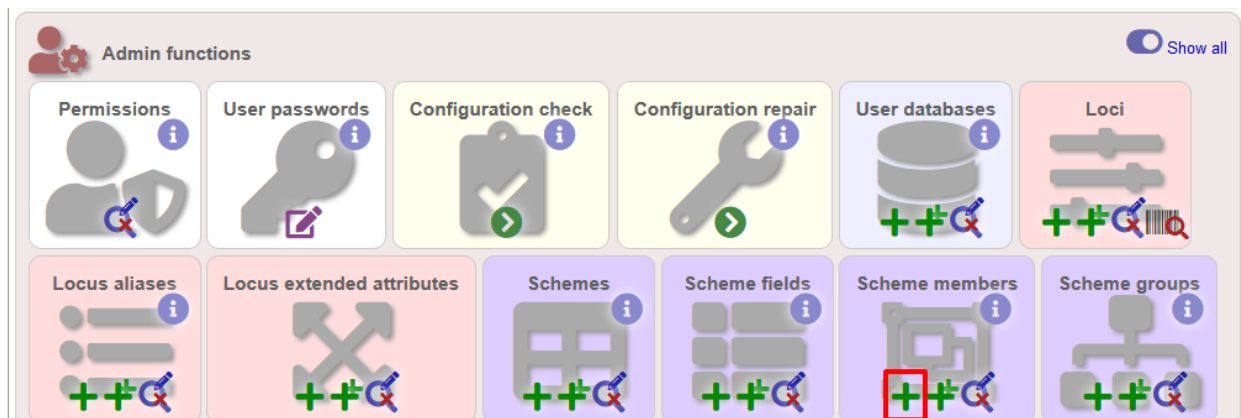
PubMed ids:

links:

(Format: URL|description)

**Action**

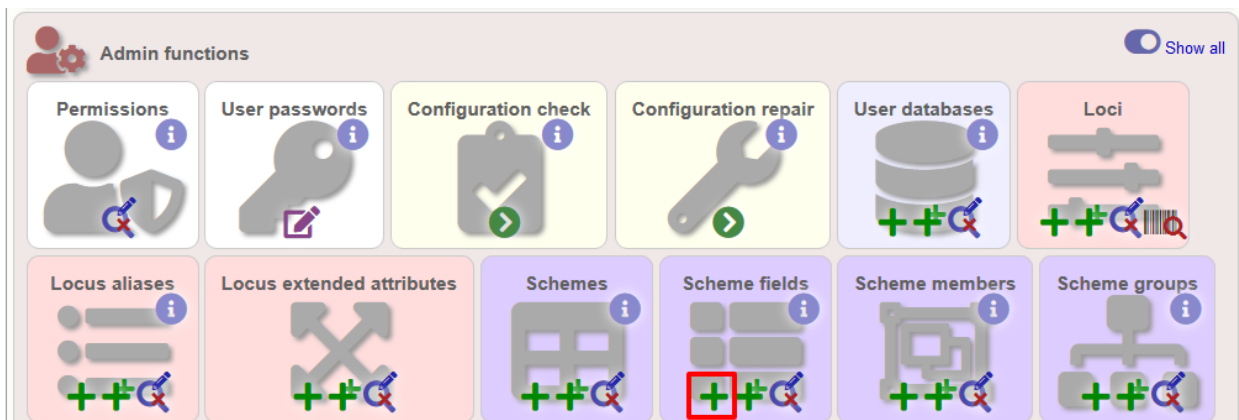
To add loci to the scheme, click the add (+) scheme members link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.







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

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



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

PubMLST Database home Curator home Contents

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

scheme id:!

MLST


field:!

ST


type:!

integer

primary key:!

☒ true ☐ false 

dropdown:!

☐ true ☒ false 

curator:!

Keith Jolley (keith)

timestamp:!


2018-06-18

description:

field order:

1

index:


☐ true ☐ false 

value regex:

Action

Reset

Submit



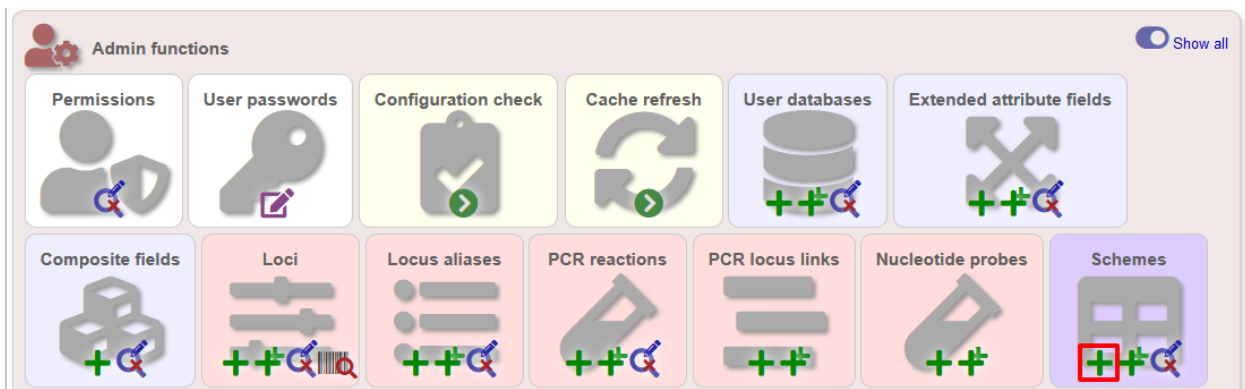
- `scheme_id` - Dropdown box of scheme names.
  - Allowed: selection from list.
- `field` - Field name.
  - Allowed: any value.
- `type` - Format for values.
  - Allowed: text/integer/date.
- `primary_key` - Set to true if field is the primary key. There can only be one primary key for a scheme.
  -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


### 5.16.2 Isolate databases

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

Click the add (+) scheme link on the curator’s interface contents page. This function is normally hidden, so you may need to click the ‘Show all’ toggle to display it.



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


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: 109


name: 

isolate display: ☒ true ☐ false 

main display: ☒ true ☐ false 

query field: ☒ true ☐ false 

query status: ☐ true ☒ false 


analysis: ☒ true ☐ false 


curator: Keith Jolley (keith)


datestamp: 2016-07-26


date entered: 2016-07-26


description:


dbase name: 


dbase host: 


dbase port: 

dbase user: 

dbase password: 

dbase id: 

display order: 

allow missing loci: ☐ true ☐ false 

flags: 

citation required

experimental

in development

unpublished

Use CTRL/SHIFT click to select or deselect values

PubMed ids:

links:

(Format: URL|description)

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\_id - Id of scheme in the sequence definition database.
  - Allowed: any integer.
- view - Restrict this scheme to only isolates contained in the specified database view. This option will only appear if the views attribute is set in the system tag. The view needs to have been defined in the database as a subset of the isolate table (usually filtered by the value of one or more of its fields).
- display\_order - Integer reflecting the display position for this scheme within the interface (optional).
  - Allowed: any integer.
- allow\_missing\_loci - Allow profile definitions to contain '0' (locus missing) or 'N' (any allele).

## 5.17 Organizing schemes into hierarchical groups

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

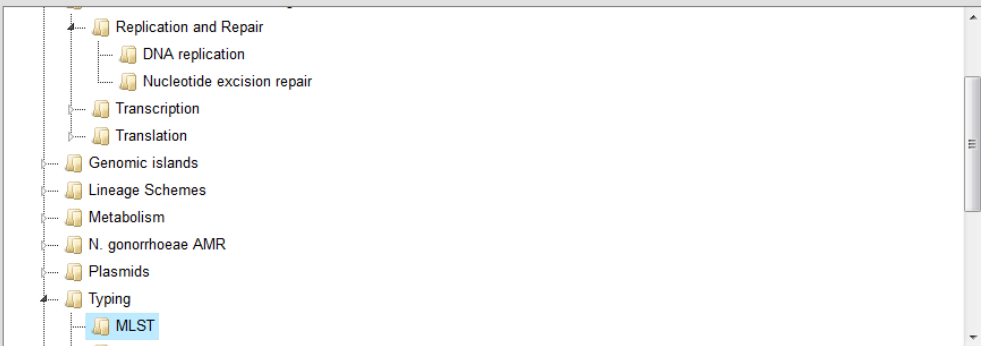
PubMLST Database home Contents

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



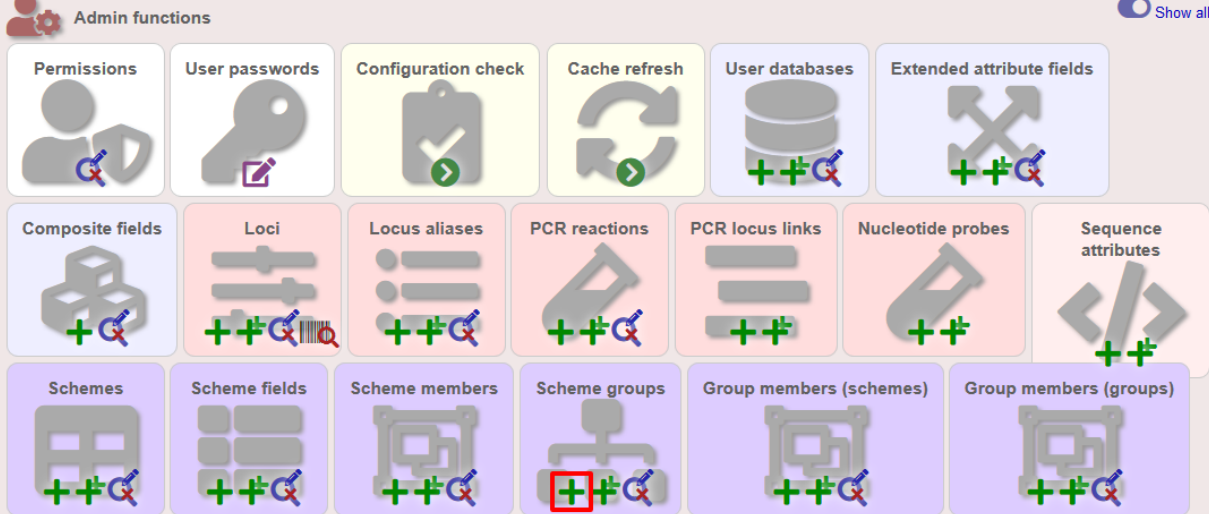
### MLST

Locus	Download	Type	Alleles	Length (setting)	Min length	Max length	Full name/product	Curator(s)	Last updated
abcZ		DNA	879	Fixed: 433 bp	433	434		O. Harrison, K. Jolley	2018-06-11
adk		DNA	651	Fixed: 465 bp	465	465		O. Harrison, K. Jolley	2018-06-11
aroE		DNA	929	Fixed: 490 bp	490	493		O. Harrison, K. Jolley	2018-06-11
fumC		DNA	976	Fixed: 465 bp	463	465		O. Harrison, K. Jolley	2018-06-11
gdh		DNA	911	Fixed: 501 bp	501	513		O. Harrison, K. Jolley	2018-06-11
pdhC		DNA	897	Fixed: 480 bp	480	503		O. Harrison, K. Jolley	2018-06-11
pgm		DNA	932	Fixed: 450 bp	447	450		O. Harrison, K. Jolley	2018-06-11

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

Admin functions Show all



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

PubMLST Database home Curator home Contents

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: 2018-06-18

description:

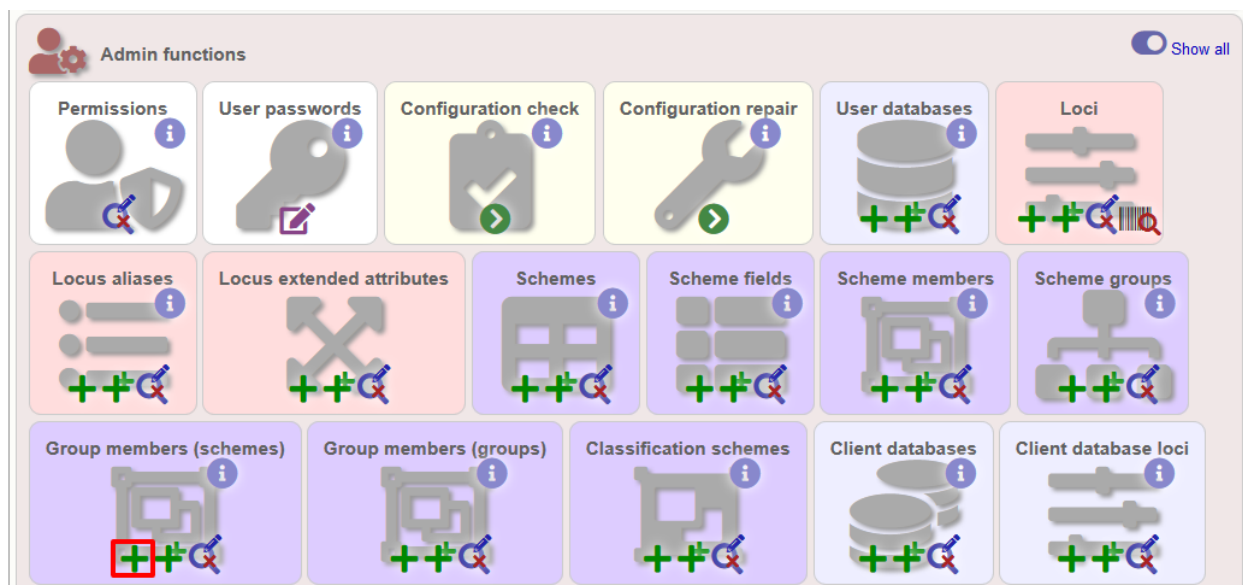
display order:

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 (+) group members (scheme) link.



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

PubMLST Database home Curator home Contents

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"/> scheme id:! <input type="text" value="MLST"/> curator:! <b>Keith Jolley (keith)</b> datestamp:! <b>2018-06-18</b>	<input type="button" value="Reset"/> <input type="button" value="Submit"/>

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

## 5.18 Setting up client databases

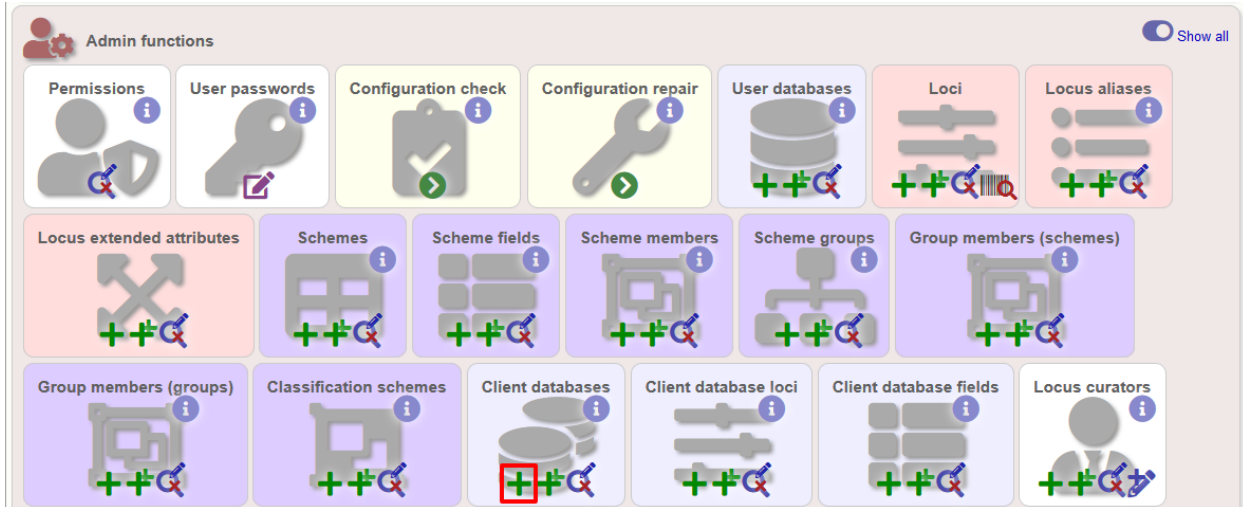
Sequence definition databases can have any number of isolate databases that connect as clients. Registering these databases allows the software to perform isolate data searches relevant to results returned by the sequence definition database, for example:

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

Multiple client databases can be queried simultaneously.

To register a client isolate database for a sequence definition database, click the add (+) client database link on the curator's interface contents page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.





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

PubMLST Database home Curator home Contents

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 (!).

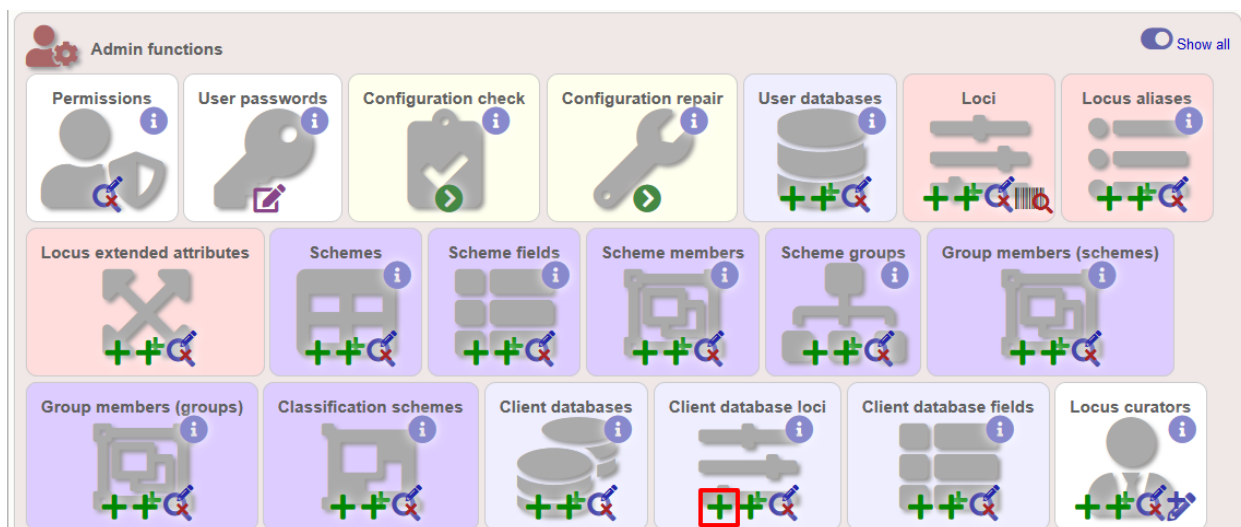
Record	Action
<p>id: 1</p> <p>name: PubMLST isolates</p> <p>description: Contains data for a collection of isolates that represent the total known diversity of <i>Neisseria</i> species. For every allelic profile in the profiles database there is at least one corresponding isolate deposited here. Any isolate may be submitted to this database and consequently it should be noted that it does not represent a population sample.</p> <p>dbase name: pubmlst_bigsdb_neisseria_isolates <small>Name of the database holding isolate data</small></p> <p>dbase config name: pubmlst_neisseria_isolates <small>Name of the database configuration</small></p> <p>curator: Keith Jolley (keith)</p> <p>timestamp: 2018-06-19</p> <p>dbase host: <small>IP address of database host</small></p> <p>dbase port: <small>Network port accepting database connections</small></p> <p>dbase user: <small></small></p> <p>dbase password: <small></small></p> <p>dbase view: isolates <small>View of isolates table to use</small></p> <p>url: /cgi-bin/bigsdb/bigsdb.p <small>Web URL to database script</small></p>	<p>Reset Submit</p>

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

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

### 5.18.1 Look up isolates with given allele

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



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

PubMLST Database home Curator home Contents

Logged in: **Keith Jolley (keith)** Log out Change password Toggle: 1

### Add new locus to client database definition

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

**Record**

client dbase id: 1) PubMLST isolates

locus: abcZ

curator: Keith Jolley (keith)

timestamp: 2018-06-19

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

**Action**

Reset Submit

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

PubMLST Database home Contents

Log in Help

### Allele information - abcZ: 5

**Provenance/meta data**

locus: abcZ

allele: 5

sequence: TTTGATACCG TTGCCGAAGG TTTGGGCGAA ATTGCGGATT TATTGCGCCG TTATCATCAT GTCAGCCATG AGTTGGAAAA TGGTTCGAGT  
GAGGCTTTGT TGAAAGAGCT TAACGAATTG CAACTTGAAA TCGAAGCGAA GGACGGCTGG AAGCTGGATG CGGCAGTCAA GCAGACTTTG  
GGTGAACCTG GTTTGCCAGA AAACGAAAAA ATCGGCAACC TCTCGGCGCG ACAGAAAAAG CGTGTGCCC TAGCGCAGGC TTGGGTGCAG  
AAGCCTGATG TATTGCTGCT GGACGAACCG ACCAACCATT TGGACATTGA CGCGATTATT TGGCTGGAAA ATCTGCTTAA AGCGTTTGAA  
GGCAGCCTGG TTGTGATTAC CCACGACCGC CGTTTTTTGG ACAATATGCG CACGCGCATC GTCGAACCTG ATC

length: 433

status: Sanger trace checked

date entered: 2001-02-07

timestamp: 2009-11-11

sender: Keith Jolley, University of Oxford, UK

curator: Man-Suen Chan, University of Oxford

**Profiles containing this allele**

MLST: 262 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. 859 isolates

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

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

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


PubMLST Database home Contents

Log in Help Toggle

### 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


sender: Paula Kriz, Paula Kriz and Keith Jolley  
 curator: Keith Jolley, University of Oxford, UK  
 date entered: 2001-02-07  
 datestamp: 2009-11-11








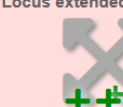
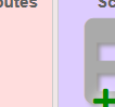





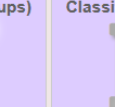




 Client database

**PubMLST isolates:** Contains data for a collection of isolates that represent the total known diversity of *Neisseria* species. For every allelic profile in the 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. [5433 isolates](#)

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



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

 Admin functions Show all

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

PubMLST Database home Curator home Contents

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

**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:!  
locus:!  
isolate field:!  
curator:!  
datestamp:!  
allele query:


1) PubMLST isolates  
penA  
penicillin\_range  
Keith Jolley (keith)  
2018-06-19  
☒ 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.


5.18. Setting up client databases

79


PubMLST Database home Contents

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


## Allele information - penA: 9

 Provenance/meta data


locus: penA  
 allele: 9  
 sequence: GACGGCGTTT TGCTGCCGGT CAGCTTTGAA AAACAGGCGG TTGCGCCGCA AGGCAACGT ATATTTAAAG CATCGACCGC ACGTCAGGTG  
 CGTGAGTTGA TGGTTTCTGT AACCGAACCT GGCGGTACGG GTACGGCGGG TCGGGTAGAT GGTTCGACG TCGCGCAAA AACCGGTACG  
 GCGCGTAAGT TGGTTAACGG TCGTTACGTC GATTACAAAC ACGTTGCCAC TTTCATCGGT TTTGCCCGG CTAATAATCC GCGTGTGATT  
 GTGGCGGTAA CCATTGACGA GCCGACTGCA AACGGTTACT ACGCGGCGCT AGTGACAGGT CCGGTCTTCA AACAAAGTTAT GGGCGGTAGC  
 CTGAACATCT TGGCGGTTTC TCCGACCAA CCTCTGACCA AT  
 length: 402  
 status: Sanger trace checked  
 date entered: 2006-09-04  
 timestamp: 2006-09-04  
 sender: Muhamed-Kheir Taha, National Reference Center for Meningococci, Institut Pasteur, France  
 curator: Keith Jolley, University of Oxford, 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, Lamberts L, Levenet I, Musilek M, Paragi M, Saguer 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. [494 isolates](#)

 Linked data

penicillin\_range: >0.06 - 1 (intermediate) [n=92] [PubMLST isolates](#)

## 5.19 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.

### 5.19.1 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 = $results->{'scheme'}->{1}->{'ST'} // 'ND';
append_html("<ul><li>Pl.$alleles{'PorA_VR1'}, $alleles{'PorA_VR2'}; $alleles{'FetA_VR'}
→ ' '; ST-$st </li></ul>");

#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.

```

(continues on next page)

(continued from previous page)

```

#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>");
}

```

### 5.19.1.1 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).

### 5.19.1.2 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](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 `<li>` and `</li>` tags) that will appear in the ‘Query database’ section of the contents page.

### 5.19.1.3 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`.

## 5.20 Workflow for setting up a MLST scheme

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

### Seqdef database

1. Create appropriate loci
2. Create new scheme ‘MLST’



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

### Isolate database

1. Create the same loci with the following additional parameters (example locus 'atpD')
  - dbase\_name: seqdef\_db
  - dbase\_id: atpD
  - url: something like /cgi-bin/bigsdb/bigsdb.pl?db=seqdef\_db&page=alleleInfo&locus=atpD&allele\_id=[?]
2. Create scheme 'MLST' with:
  - dbase\_name: seqdef\_db
  - dbase\_id: 1 (or whatever the id of your seqdef scheme is)
3. Add scheme\_field ST as before
4. Add loci as scheme\_members

## 5.21 Automated assignment of scheme profiles

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

The script is run as follows:

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

A full list of options can be found by typing:

```
define_profiles.pl --help
NAME
    define_profiles.pl - Define scheme profiles found in isolate database

SYNOPSIS
    define_profiles.pl --database NAME --scheme SCHEME_ID [options]

OPTIONS

--cache
    Update scheme field cache in isolate database.

--database NAME
    Database configuration name.

--help
    This help page.

--exclude_isolates LIST
    Comma-separated list of isolate ids to ignore.
```

(continues on next page)

(continued from previous page)

```

--exclude_projects LIST
    Comma-separated list of projects whose isolates will be excluded.

--ignore_multiple_hits
    Set allele designation to 'N' if there are multiple designations set for
    a locus. The default is to use the lowest allele value in the profile
    definition.

--isolates LIST
    Comma-separated list of isolate ids to scan (ignored if -p used).

--isolate_list_file FILE
    File containing list of isolate ids (ignored if -i or -p used).

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

--max ID
    Maximum isolate id.

--min ID
    Minimum isolate id.

--min_size SIZE
    Minimum size of seqbin (bp) - limit search to isolates with at least this
    much sequence.

--missing NUMBER
    Set the number of loci that are allowed to be missing in the profile. If
    the remote scheme does not allow missing loci then this number will be set
    to 0. Default=0.

--projects LIST
    Comma-separated list of project isolates to scan.

--scheme SCHEME_ID
    Scheme id number.

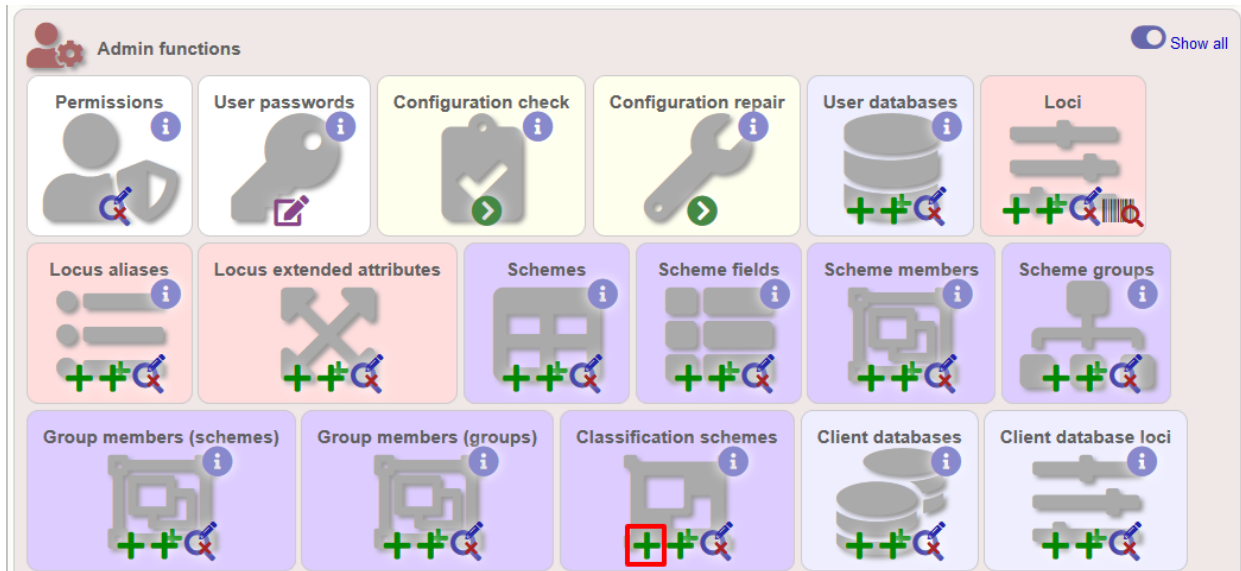
```

## 5.22 Scheme profile clustering - setting up classification schemes

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

### 5.22.1 Defining classification scheme in sequence definition database

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



Select the underlying scheme and enter a name for the classification scheme, the number of mismatches allowed in order to include a scheme profile in a group, and a description. An example name for such a scheme could be 'Nm\_cgc\_25' indicating that this is a classification scheme for *Neisseria meningitidis* core genome cluster with a threshold of 25 mismatches.

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

$$(\text{number of common loci} \times (\text{total loci} - \text{defined threshold})) / \text{total loci}$$

rather than

$$\text{total loci} - \text{defined threshold}$$

when an absolute threshold is used.

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

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

Press 'Submit' to create the classification scheme.

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith) Log out Change password Toggle: [i] [≡]

## Add new classification scheme

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

**Record**

id: 2

scheme id: N. meningitidis cgMLST v1.0

name: Nm\_cgc\_25

inclusion threshold: 25 Maximum number of different alleles allowed between profile and at least one group member profile.

use relative threshold: ☐ true ☒ false Calculate threshold using ratio of shared/present in both profiles in pairwise comparison.

status: experimental

curator: Keith Jolley (keith)

timestamp: 2018-06-19

description: Single linkage clustering with each group member having fewer than 25 allelic mismatches to at least one other member of the group. Missing loci are ignored in comparisons.

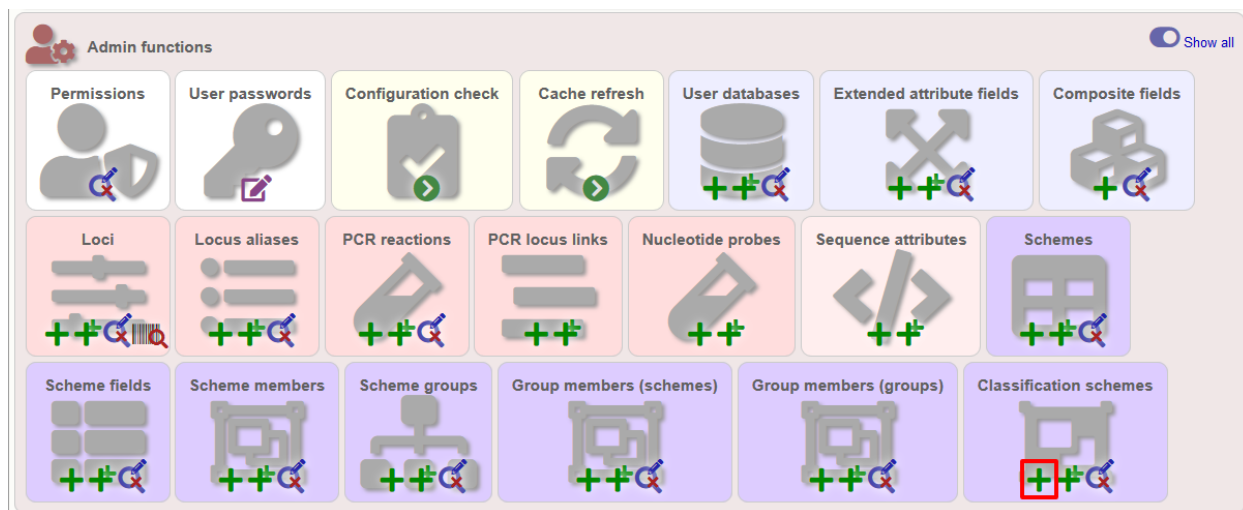
display order:

**Action**

Reset Submit

### 5.22.2 Defining classification scheme in isolate database

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



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

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

PubMLST Database home Curator home Contents

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

### Add new classification scheme

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

**Record**

id: 2

scheme id: N. meningitidis cgMLST v1.0

name: Nm\_cgc\_25

inclusion threshold: 25 Maximum number of different alleles allowed between profile and at least one group member profile.

use relative threshold: ☐ true ☒ false Calculate threshold using ratio of shared/present in both profiles in pairwise comparison.

status: experimental

curator: Keith Jolley (keith)

timestamp: 2018-06-19


description: Single linkage clustering with each group member having fewer than 25 allelic mismatches to at least one other member of the group. Missing loci are ignored in comparisons.

seqdef cscheme id: 2 cscheme\_id number defined in seqdef database

display order:

**Action**

Reset Submit



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

### 5.22.3 Clustering

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

Currently only single-linkage clustering is supported.

The script is run as follows from the command line:

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

A full list of options can be found by typing:

```
cluster.pl --help
NAME
    cluster.pl - Cluster cgMLST profiles using classification groups.

SYNOPSIS
    cluster.pl --database NAME --cscheme_id SCHEME_ID [options]

OPTIONS
--cscheme CLASSIFICATION_SCHEME_ID
    Classification scheme id number.

--database NAME
    Database configuration name.

--help
    This help page.

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

**Note:** Note that for classification schemes to be accessible within the isolate database, *scheme cache tables* must be

generated and kept up-to-date.

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

disease: invasive (unspecified/other)      Biological Standards and Controls,  
 epidemiology: endemic      Potters Bar, UK

**Publications (2)**

- Feavers IM, Gray SJ, Urwin R, Russell JE, Bygraves JA, Kaczmarek EB, Maiden MC (1999). Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak. *J Clin Microbiol* 37:3883-7 [14 isolates](#)
- Jolley KA, Hill DM, Bratcher HB, Harrison OB, Feavers IM, Parkhill J, Maiden MC (2012). Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid Web-based analysis methods. *J Clin Microbiol* 50:3046-53 [14 isolates](#)

**Sequence bin**

contigs: 259	mean length: 8,245 bp	N90 contig number: 63	N95 length (L95): 4,593
total length: 2,135,447 bp	N50 contig number: 18	N90 length (L90): 8,066	loci tagged: 2,180
max length: 130,716 bp	N50 length (L50): 38,364	N95 contig number: 79	detailed breakdown: <a href="#">Display</a>

**Similar isolates (determined by classification schemes)**

Experimental schemes are subject to change and are not a stable part of the nomenclature.

Classification scheme	Underlying scheme	Clustering method	Mismatch threshold	Status	Group
Nm_cgc_200	N. meningitidis cgMLST v1.0	Single-linkage	200	experimental	group: 17 (841 isolates)
Nm_cgc_100	N. meningitidis cgMLST v1.0	Single-linkage	100	experimental	group: 38 (588 isolates)
Nm_cgc_50	N. meningitidis cgMLST v1.0	Single-linkage	50	experimental	group: 45 (4 isolates)
Nm_cgc_25	N. meningitidis cgMLST v1.0	Single-linkage	25	experimental	group: 45 (4 isolates)

**Schemes and loci**

☒ All loci      Navigate and select schemes within tree to display allele designations  
☐ Capsule  
☐ Genetic Information Processing

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

PubMLST   Database home   Contents  
 Logged in: [Keith Jolley \(keith\)](#)   Log out   Change password   Help   Toggle

**Search or browse *Neisseria* PubMLST database**

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields      Allele designations/scheme fields  
 id   =   Enter value...   +   Nm\_cgc\_25 group   =   45   +

Display/sort options      Action  
 Order by: id   ascending   Reset   Submit  
 Display: 25 records per page

4 records returned. Click the hyperlinks for detailed information.

Your projects

Select project...   Add these records

Isolate fields										MLST		Finetyping antigens		
id	isolate	aliases	country	year	disease	species	serogroup	genogroup	capsule group	ST	clonal complex	PorA VR1	PorA VR2	FetA VR
662	2837	M97/252508	UK	1997	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	C		C	50	ST-11 complex	5-1	10-4	F3-6
670	2840	M97/252535	UK	1997	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	C		C	50	ST-11 complex	5-1	10-4	F3-6
671	2844	M97/252781	UK	1997	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	C		C	50	ST-11 complex	5-1	10-4	F3-6
672	2847	M97/252943	UK	1997	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	C		C	50	ST-11 complex	5-1	10-4	F3-6

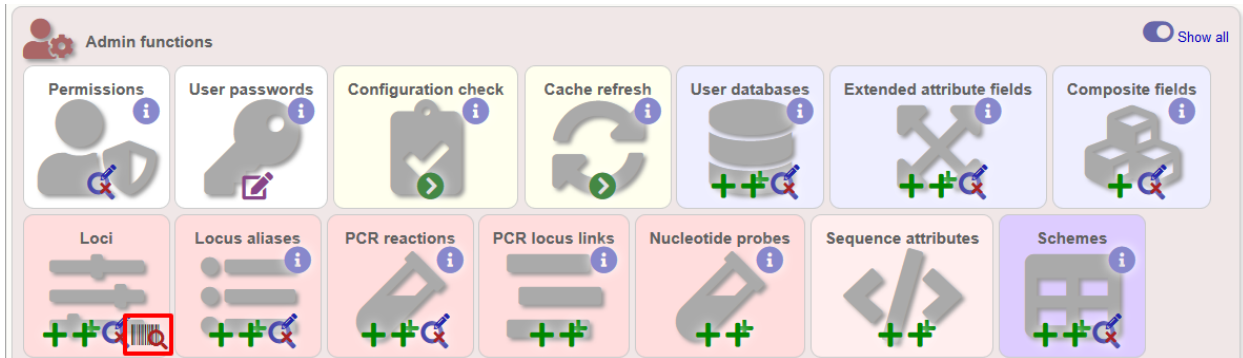
**Analysis tools:**

Breakdown: Fields   Two Field   Combinations   Polymorphic sites   Schemes/alleles   Publications   Sequence bin   Tag status  
 Analysis: BURST   Codons   Presence/Absence   Genome Comparator   BLAST   rMLST species id  
 Export: Dataset   Contigs   Sequences  
 Third party: GrapeTree   iTOL   PhyloViz   Microreact

## 5.23 Defining new loci based on annotated reference genome

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

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



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

PubMLST Database home Curator home Contents

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: <input type="text" value="AM421808"/>	<input checked="" type="radio"/> locus tag <input type="radio"/> gene name	<input type="button" value="Submit"/>

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

PubMLST Database home Curator home Contents

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: AM421808 ☒ 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	1326
NMC0011		putative integral membrane protein	840
NMC0012		putative lipoprotein	1167

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.



PubMLST Database home Curator home Contents

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: AM421808 ☒ locus tag ☐ gene name Submit

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	1326
NMC0011		putative integral membrane protein	840
NMC0012		putative lipoprotein	1167

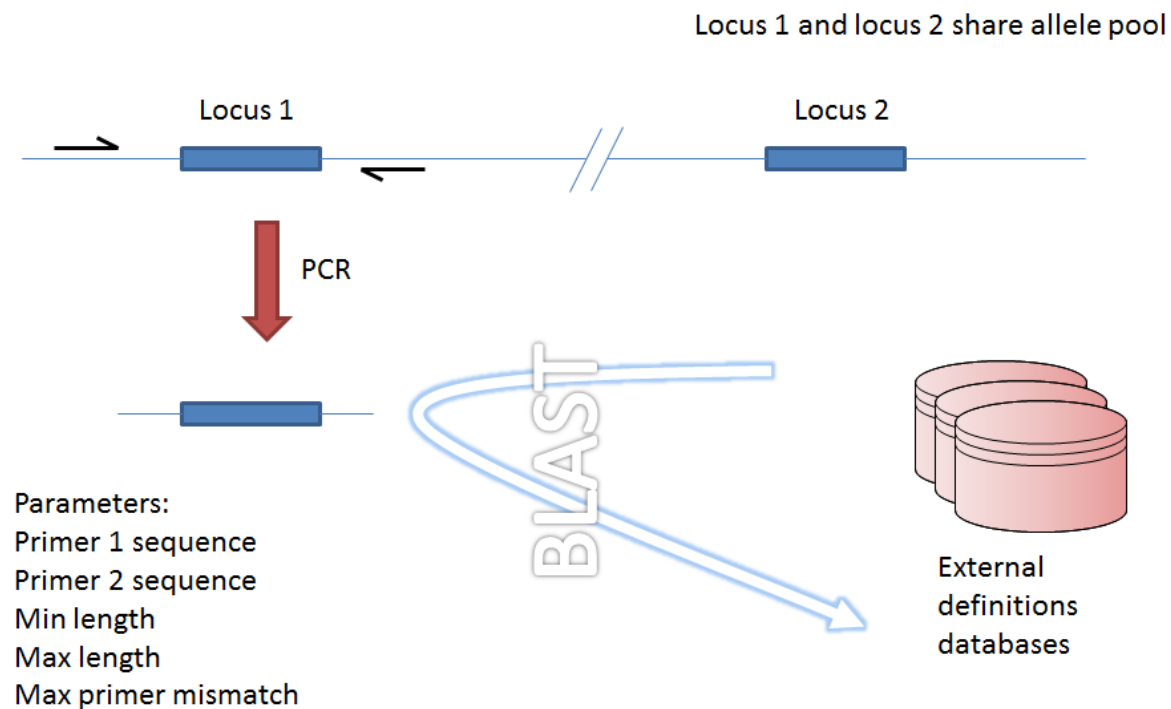
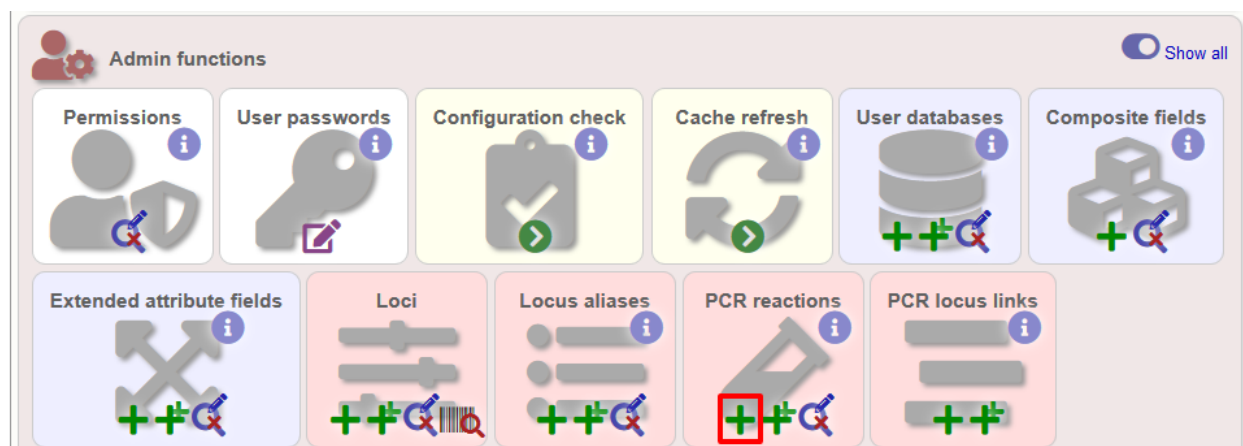
## 5.24 Genome filtering

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

### 5.24.1 Filtering by *in silico* PCR



Provided a locus can be predicted to be specifically 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. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.

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

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


PubMLST Database home Curator home Contents

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:! 2 

description:!

primer1:!



primer2:!

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


datestamp:! **2018-06-19**

min length:  Minimum length of product to return

max length:  Maximum length of product to return

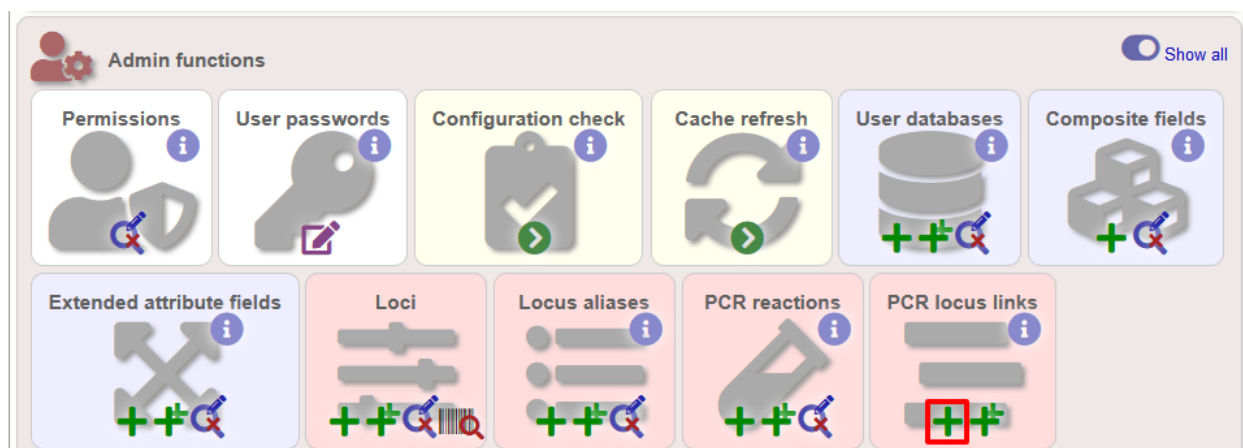
max primer mismatch:   Maximum sequence mismatch per primer

**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.



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

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

### 5.24.2 Filtering by *in silico* hybridization

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

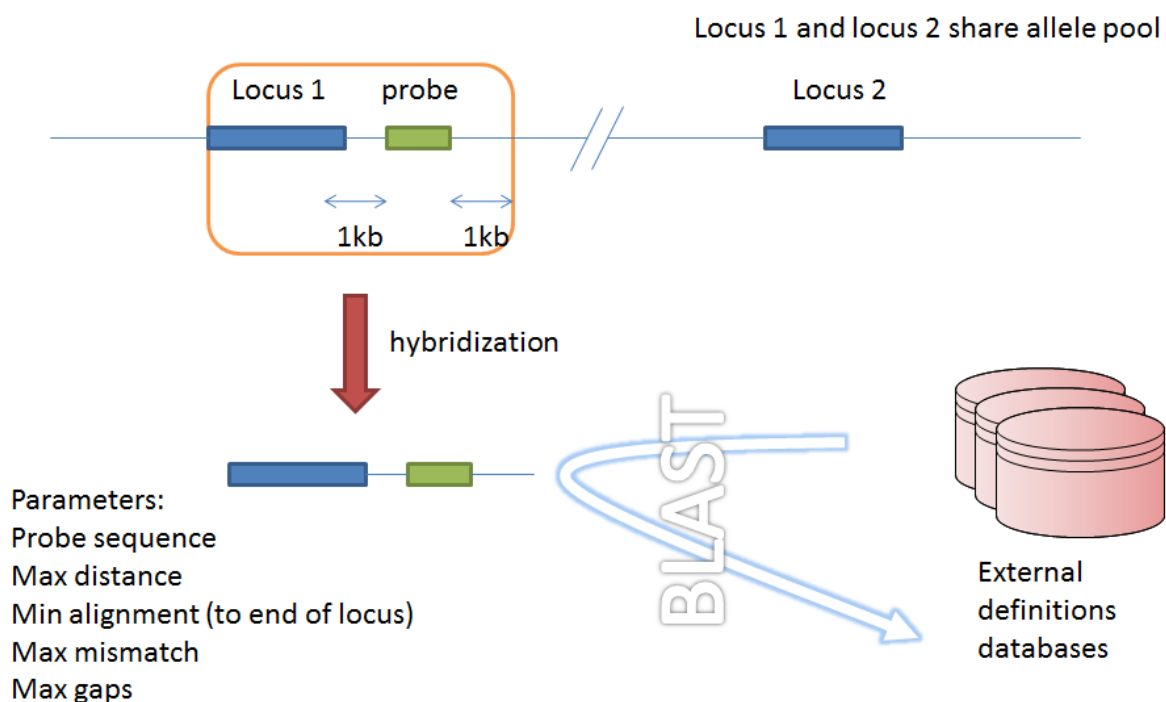
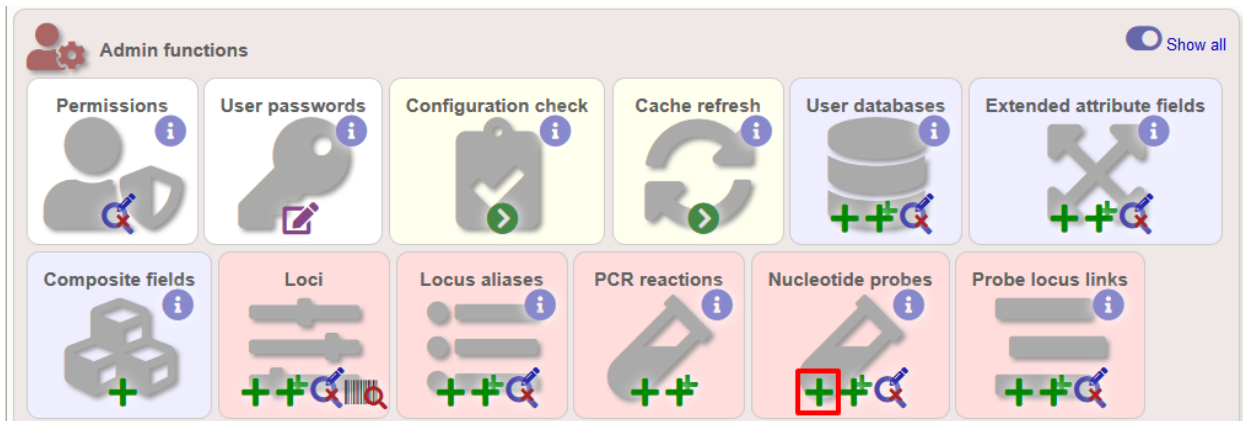
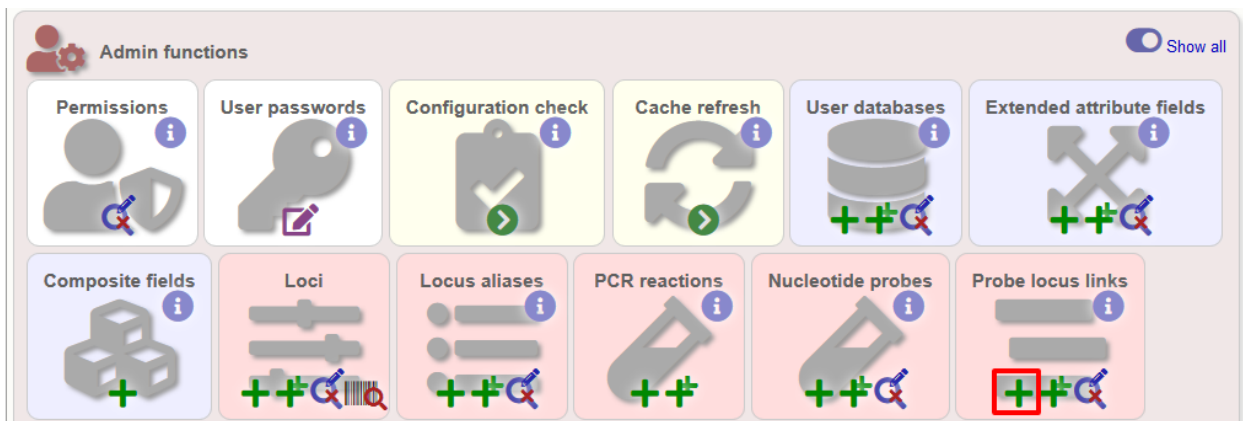


Fig. 2: Filtering by *in silico* hybridization

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



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



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

- probe\_id - Dropdown list of probe names.
  - Allowed: selection from list.
- locus - Dropdown list of loci.
  - Allowed: selection from list.
- max\_distance - Minimum distance of probe from end of locus.
  - Allowed: any positive integer.
- min\_alignment - Minimum length of alignment allowed.
  - Allowed: any positive integer.
- max\_mismatch - Maximum number of mismatches allowed in alignment.
  - Allowed: any positive integer.
- max\_gaps - Maximum number of gaps allowed in alignment.
  - Allowed: any positive integer.

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

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

## 5.25 Setting locus genome positions

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

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



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

PubMLST Database home Curator home Contents

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

### Isolate query/update

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields  
 isolate = mc58 + [i]

Display/sort options  
 Order by: id ascending  
 Display: 25 records per page [i]

Action  
 Reset Submit

1 record returned. Click the hyperlink for detailed information.

Delete Tag scanning Projects Your projects  
 Delete ALL Scan Select project... Link Select project... Add these records

Delete	Update	Sequence bin	New version	Isolate fields [i]										MLST		Finotyping antigens		
				id	isolate	aliases	country	year	disease	species	serogroup	genogroup	capsule group	ST	clonal complex	PorA VR1	PorA VR2	FetA VR
				<a href="#">240</a>	MC58	Z7176	UK	1983		Neisseria meningitidis	B		B	74	ST-32 complex	7	16-2	F1-5

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

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith) Log out Change password Help Toggle: 1

### Full information on isolate MC58 (id:240)

[Delete record](#) [Update record](#) [Sequence bin](#) [New version](#) [Sequence tags](#)  
[Delete](#) [Update](#) [Upload contigs](#) [Create](#) [Scan](#)

**Provenance/meta data**

id: 240  
 isolate: MC58  
 alias: Z7176  
 strain designation: B: P1.7,16-2: F1-5: ST-74 (cc32)  
 country: UK  
 continent: Europe  
 year: 1983  
 species: Neisseria meningitidis  
 serogroup: B  
 capsule group: B

serotype: 15  
 sero subtype: P1.7,16-2  
 comments: Genome sequenced by TIGR.  
 sender: Mumtaz Virji, Dept. Pathology and Microbiology, University of Bristol  
 curator: Nina Billows, University of Oxford (E-mail: [nina.billows@some.ox.ac.uk](mailto:nina.billows@some.ox.ac.uk))  
 update history: [313 updates](#) [show details](#)  
 date entered: 2001-05-11  
 datestamp: 2018-06-04

**Publications (6)**

**Sequence bin**

contigs: 1  
 length: 2272360 bp  
 loci tagged: 2,226  
 detailed breakdown: [Display](#)

Click the ‘Renumber’ button:

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith) Log out Change password Help Toggle: 1

### Sequence bin for MC58

**Contig summary statistics**

Contigs: 1  
 Length: 2,272,360

- Download sequences (FASTA format)
- Download sequences with annotations (EMBL format)

Sequence	Sequencing method	Original designation	Length	Comments	Locus	Start	End	Direction	EMBL format	Artemis	Renumber
1	Sanger		2272360	whole genome	NEIS2139	7	498	←	EMBL	Artemis	Renumber
					NEIS2140	502	897	←			
					NEIS2141	918	2312	←			
					NEIS2142	2517	3161	←			
					NEIS2143	3158	3511	←			
					NEIS2144	3635	4117	→			
					NEIS2145	4311	4961	→			
					NEIS2146	4958	5875	→			
					NEIS2147	5936	6214	→			
					NEIS2148 (pgk)	6281	7492	←			
					NEIS2149	7573	8826	←			
					tRNA-lys	9197	9272	←			
					NEIS2150	9346	10317	←			
					NEIS2151	10350	10811	←			

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 home Curator home Contents

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
NEIS2139		7
NEIS2140		502
NEIS2141		918
NEIS2142		2517
NEIS2143		3158
NEIS2144		3635
NEIS2145		4311
NEIS2146		4958
NEIS2147		5936
NEIS2148		6281
NEIS2149		7573
tRNA-lys		9197
NEIS2150		9346
NEIS2151		10350
NEIS2152		10840
NEIS2153		12174
NEIS2979		13848
NEIS0001		15221
NEIS0210		17229
NEISp0210		17232
pilS		18127

## 5.26 Defining composite fields

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

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

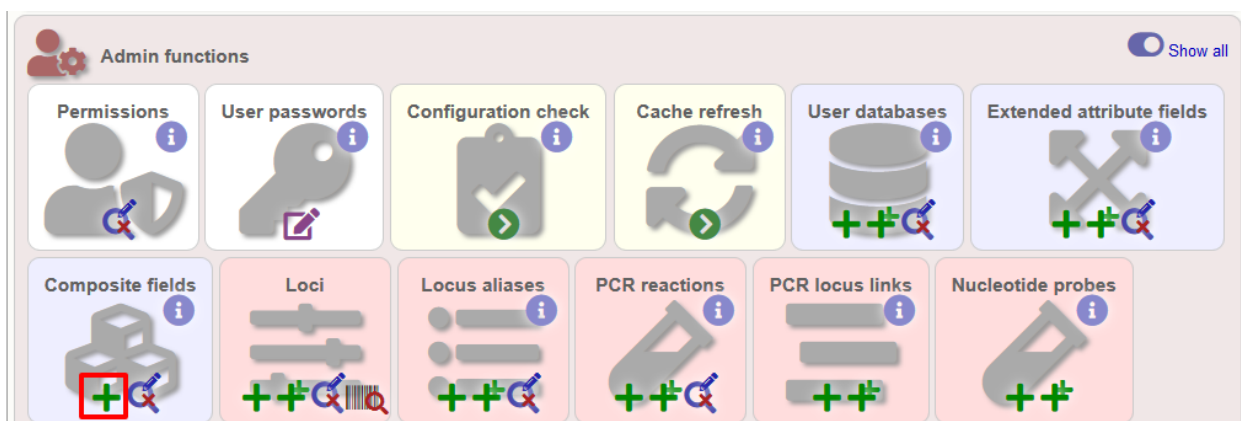
[serogroup]: P1.[PorA\_VR1],[PorA\_VR2]: [FetA\_VR]: ST-[ST] ([clonal\_complex])

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

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

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





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

PubMLST Database home Curator home Contents

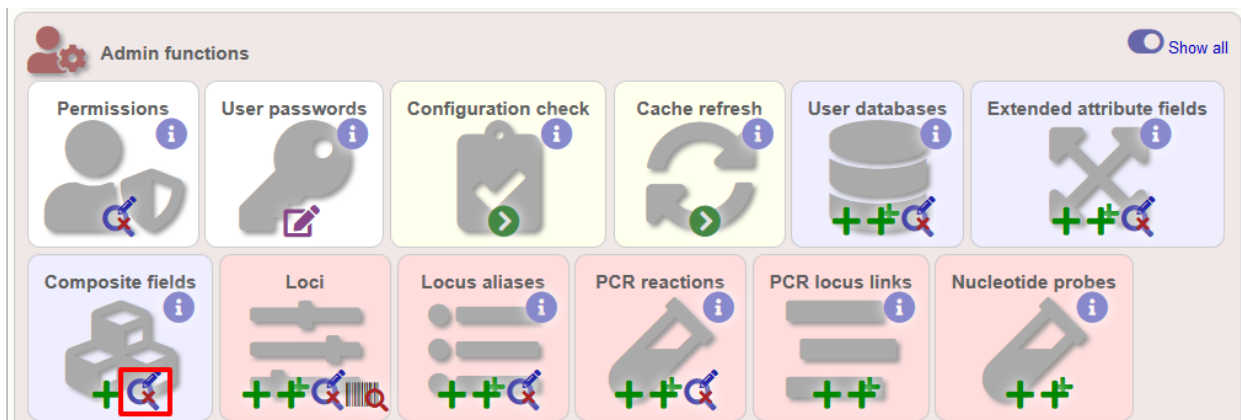
Logged in: Keith Jolley (keith) Log out Change password Toggle: i

### Add new composite field

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

Record	Action
<p>id: strain_designation name of the field as it will appear in the web interface</p> <p>position after: isolate field present in the isolate table</p> <p>main display: <input type="radio"/> true <input checked="" type="radio"/> false Sets whether to display field in isolate query results table (can be overridden by user preference).</p> <p>curator: Keith Jolley (keith)</p> <p>timestamp: 2018-06-19</p>	<p>Reset Submit</p>

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



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

PubMLST Database home Curator home Contents

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
		strain_designation	isolate	false	[capsule_group]: P1.[PorA_VR1].[PorA_VR2]: [FetA_VR]: ST-[scheme 1:ST] ([scheme 1:clonal_complex])	ND: P1.ND,ND: F-ND: ST-ND (-)

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

PubMLST Database home Curator home Contents

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

## Update composite field - strain\_designation

Position/display

position after: isolate

main display: ☐ true ☒ false

Update

field	empty value	regex	curator	timestamp	delete	edit	move
capsule_group [isolate field]	ND		Keith Jolley	2017-08-15			
: P1.			Keith Jolley	2009-11-12			
PorA_VR1 [locus]	ND		Keith Jolley	2009-11-12			
PorA_VR2 [locus]	ND		Keith Jolley	2009-11-12			
FetA_VR [locus]	F-ND		Keith Jolley	2009-11-12			
: ST-			Keith Jolley	2009-11-12			
ST [MLST field]	ND		Keith Jolley	2009-11-12			
(			Keith Jolley	2009-11-12			
clonal_complex [MLST field]	-	s/ST- (\S+) complex.* /cc\$1/	Keith Jolley	2009-11-12			
)			Keith Jolley	2009-11-12			

Add new field:

text field:  +

isolate field:  +

locus field:  +

scheme field:  +

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

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

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

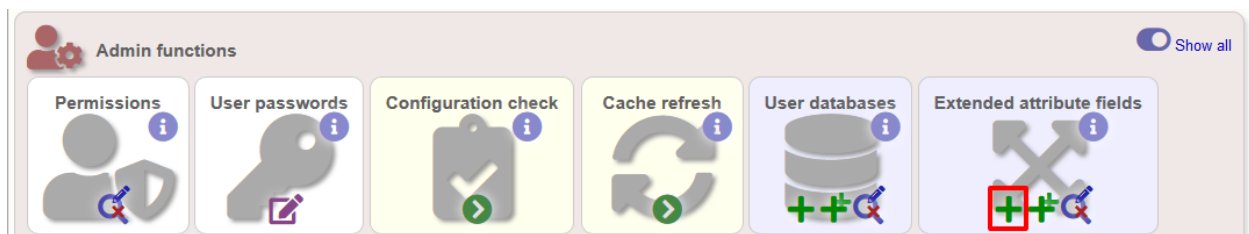
This will convert 'ST-4 complex/subgroup IV' to 'cc4'.

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

## 5.27 Extended provenance attributes (lookup tables)

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

To set up such an extended attribute, click the add (+) isolate field extended attributes link on the curator's main page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.

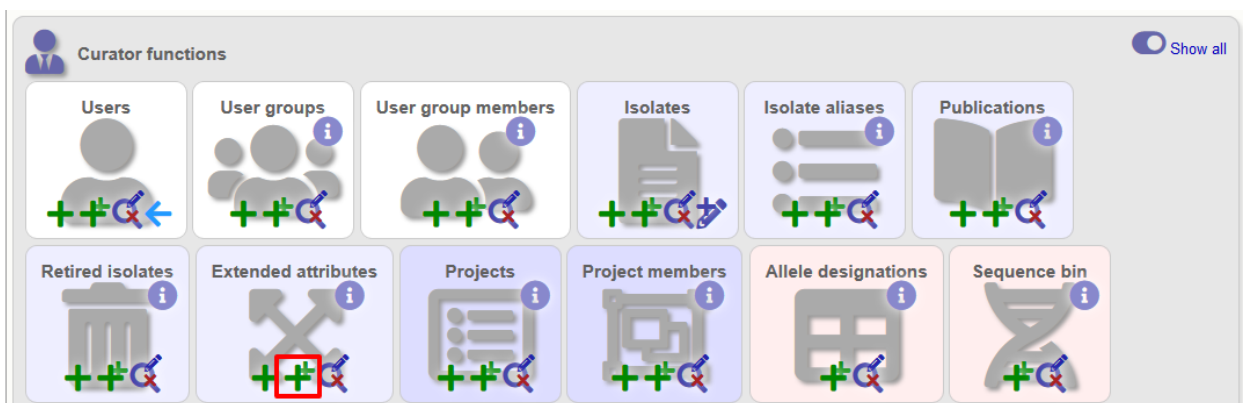


Fill in the web form with appropriate values. Required fields have an exclamation mark (!) next to them:

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

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

The easiest way to populate the lookup table is to do a batch update copied from a spreadsheet. Click the batch add (++) isolate field extended attribute values link on the curator's main page (this link will only appear once an extended attribute has been defined). This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



Download the Excel template:

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith), Log out | Change password Toggle: i

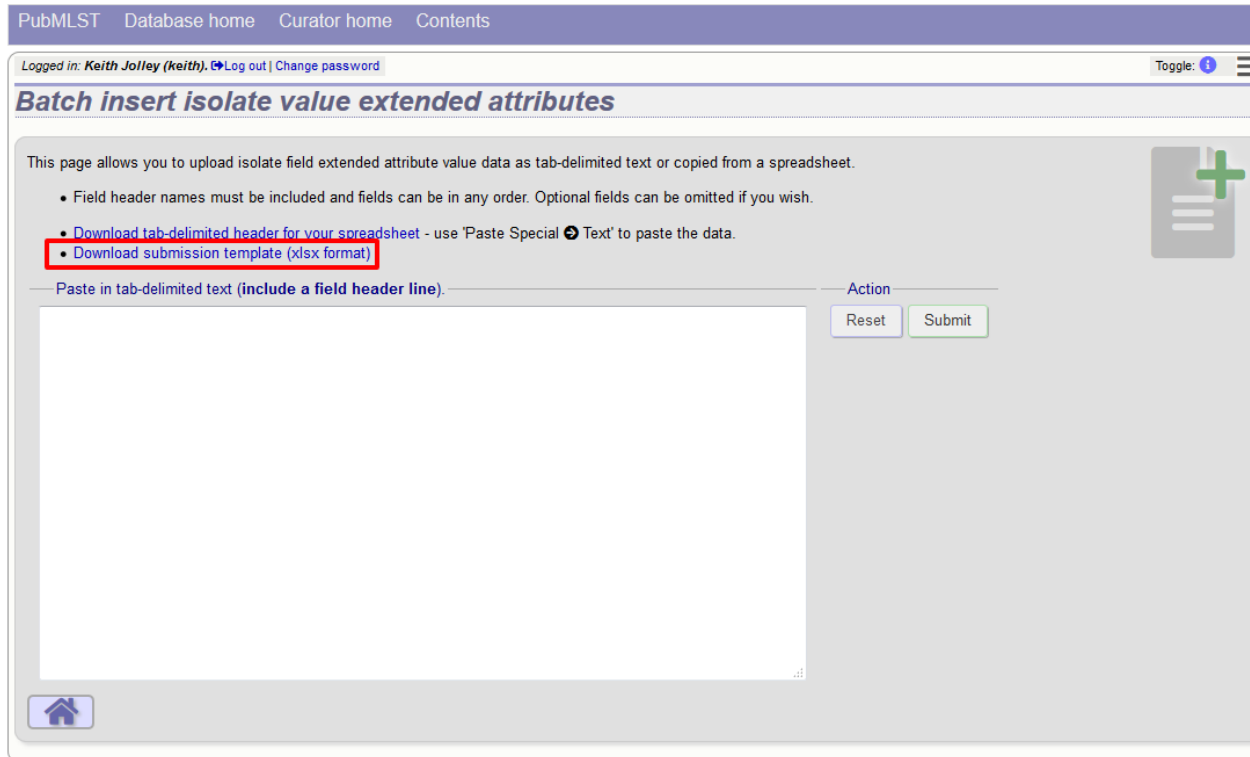
### Batch insert isolate value extended attributes

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

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- Download tab-delimited header for your spreadsheet - use 'Paste Special > Text' to paste the data.
- Download submission template (xlsx format)

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

Action  
Reset Submit



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

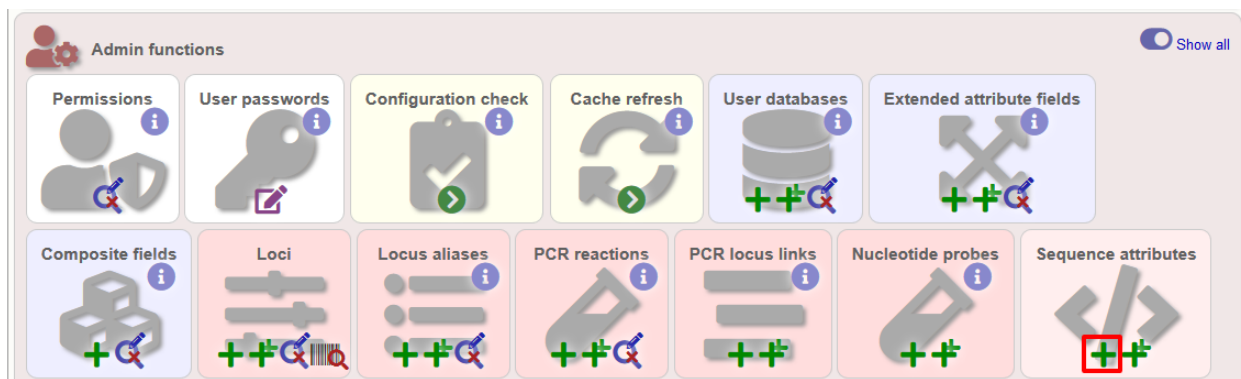
isolate_field	attribute	field_value	value
country	continent	Afghanistan	Asia
country	continent	Albania	Europe
country	continent	Algeria	Africa
country	continent	Andorra	Europe
country	continent	Angola	Africa

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

## 5.28 Sequence bin attributes

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

To set up new attributes, click the add (+) 'sequence attributes' link on the isolate database curator's index page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



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

PubMLST Database home Curator home Contents

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	Action
<div style="border: 2px solid red; padding: 5px;">           key: <input type="text" value="read_length"/>            type: <input type="text" value="integer"/> </div> curator: <b>Keith Jolley (keith)</b> timestamp: <b>2018-06-19</b> description: <input type="text"/>	<input type="button" value="Reset"/> <input type="button" value="Submit"/>

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

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (Keith) Log out Change password Toggle: 1

## 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: !

sender: ! Select sender ...

method:

run id:

assembly id:

read length:

Options

☐ Don't insert sequences shorter than 25 bps.

Link to experiment:

Alternatively upload FASTA file or enter Genbank accession

Select FASTA file:

No file selected.

Action

## 5.29 Checking external database configuration settings

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



The software will check that required helper applications are installed and executable and, in isolate databases, test every locus and scheme external database to check for connectivity and that data can be retrieved. By default, only loci which have an issue will be displayed but you can click the 'show all loci' link to display them all.

PubMLST Database home Curator home Contents

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

### Configuration check - Neisseria PubMLST

Helper applications

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

Locus databases (only showing loci with potential problems - [show all loci](#))

Locus	Database	Host	Port	Id field value	Database accessible	Sequence query	Sequences assigned
NEIS0895 (parA)	pubmlst_bigsdb_neisseria_seqdef	zoo-aberlour	5432	NEIS0895	✓	✓	✗
NEIS0903 (opaD)	pubmlst_bigsdb_neisseria_seqdef	zoo-aberlour	5432	NEIS0903	✓	✓	✗
NEIS1454	pubmlst_bigsdb_neisseria_seqdef	zoo-aberlour	5432	NEIS1454	✓	✓	✗
NEIS1551 (opaC)	pubmlst_bigsdb_neisseria_seqdef	zoo-aberlour	5432	NEIS1551	✓	✓	✗
NEIS2013	pubmlst_bigsdb_neisseria_seqdef	zoo-aberlour	5432	NEIS2013	✓	✓	✗
NEIS2538	pubmlst_bigsdb_neisseria_seqdef	zoo-aberlour	5432	NEIS2538	✓	✓	✗

Scheme databases

Scheme description	Database	Host	Port	Id	Database accessible	Profile query
MLST	pubmlst_bigsdb_neisseria_seqdef	zoo-aberlour	5432	1	✓	✓
Ribosomal MLST	bigsdb_multispecies_seqdef	zoo-aberlour	5432	1	✓	✓
rplF species	pubmlst_bigsdb_neisseria_seqdef	zoo-aberlour	5432	42	✓	✓

Any problems will be highlighted with a red X.

## 5.30 Exporting table configurations

Sometimes it is useful to transfer configurations between different databases or to export a configuration for troubleshooting. Data from most of the tables can be exported in tab-delimited text format suitable for batch uploading. For example, to export scheme configuration data, click the query link (Update or delete) next to schemes in the curator's interface. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.

Admin functions Show all

Permissions

User passwords

Configuration check

Cache refresh

User databases

Extended attribute fields

Composite fields

Loci

Locus aliases

PCR reactions

PCR locus links

Nucleotide probes

Sequence attributes

Schemes

Scheme fields

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



PubMLST Database home Curator home Contents

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

### Query schemes for Neisseria PubMLST database

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

**Search criteria**

id = + i

**Display**

Order by: id ascending

Display: 25 records per page i

**Filter query by**

isolate display: i

main display: i

query field: i

query status: i

analysis: i

allow missing loci: i

curator: i

scheme: MLST i

**Action**

Reset Submit

Click the button 'Export configuration/data'.

PubMLST Database home Curator home Contents

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

### Query schemes for Neisseria PubMLST database

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

**Search criteria**

id = + i

**Display**

Order by: id ascending

Display: 25 records per page i

**Filter query by**

isolate display: i

main display: i

query field: i

query status: i

analysis: i

allow missing loci: i

curator: i

scheme: MLST i

**Action**

Reset Submit

1 record returned.

**Delete**

Delete ALL

**Database configuration**

Export configuration/data

Delete	Update	id	name	dbase name	dbase host	dbase port	dbase id	isolate display*	main display*	query field*	query status*	analysis*	display order	allow missing loci	curator	datestamp	date entered
		1	MLST	pubmlst_bigsdb_neisseria_seqdef			1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	1	<input type="checkbox"/>	Keith Jolley	2012-03-22	2009-

\* 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  main_display  query_field query_status  analysis display_
↪order allow_missing_loci  curator  datestamp  date_entered
1 MLST  pubmlst_bigsdb_neisseria_seqdef  mv_scheme_1 1 1 1 1 1 1
↪ 2 2012-03-22 2009-11-12

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

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

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

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

```
create_client_credentials.pl --help

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

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
```

(continues on next page)

(continued from previous page)

```
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).



## CHAPTER 6

### Curator's guide

Please note that links displayed within the curation interface will vary depending on database contents and the permissions of the curator. Some infrequently used links are usually hidden by default. These can be enabled by clicking the 'Show all' toggle switch.



#### 6.1 Adding new sender details

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

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



Enter the user's details in to the form.

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

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

## 6.2 Adding new allele sequence definitions

### 6.2.1 Single allele

To add a single new allele, click the sequences add (+) link on the curator's main page.



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

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

- Sanger trace checked

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

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

[PubMLST](#)
[Database home](#)
[Curator home](#)
[Contents](#)

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:

status:

sender:

curator: **Keith Jolley (keith)**

date entered: **2018-06-07**

datestamp: **2018-06-07**

type allele: ☐ true ☐ false New allele searches can be constrained to use just type alleles in comparisons

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

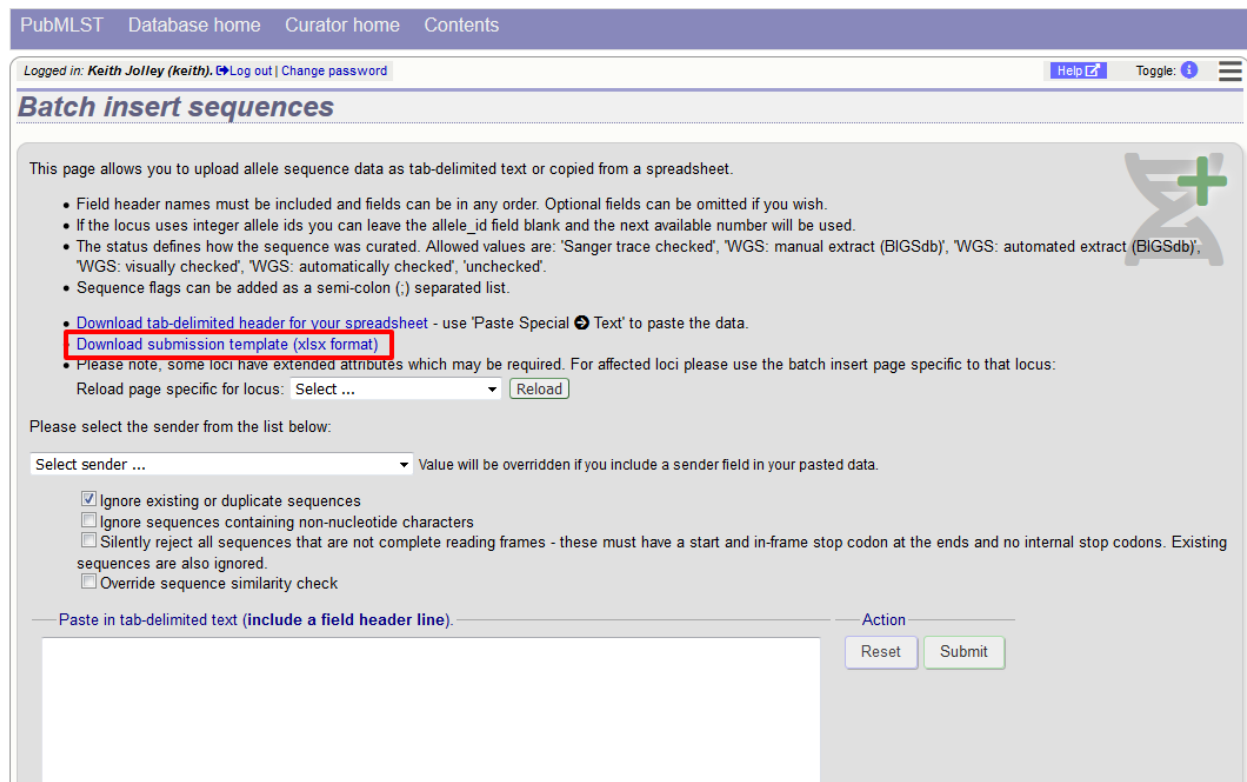
There are two methods of batch adding alleles. You can either upload a spreadsheet with all fields in tabular format, or you can upload a FASTA file provided all sequences are for the same locus and have the same status.

### 6.2.2.1 Upload using a spreadsheet

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



Download a template Excel file from the following page.



Fill in the spreadsheet. If the locus uses integer allele identifiers, the allele\_id can be left blank and the next available number will be used automatically.

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

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

Paste the entire sheet in to the web form and select the sender from the dropdown box.

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

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

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith) Log out Change password Help 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 (BIGSdb)', 'WGS: automated extract (BIGSdb)', 'WGS: visually checked', 'WGS: automatically checked', '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)

Action

Reset Submit

```
locus allele_id status sequence
abcZ WGS: automated extract (BIGSdb)
TTTGATACTGTTGCCGAAGSTTTGGSCGAAATTCGCGATTITATTCGCCGCTTATCATGTGAGCCATGAGTTGGAAAAATG
GTTCGAGTGAAGGCTTTGTTGAAAGAACTCAACGAATTGCAACTTGCAATCGAAGCGAAGGACGGCTGGAAATGGATGCGGC
AGTCAAGCAGACTTTGGGGGAACCTCGGTTTGGCGAAAAATGAAAAAATCGGCAACCTTTCCGGCGGTGAGAAAAAGCGGTC
GCTTGGCGCAGGCTTTGGGTGCAAAAGCCCGACGTAATTGCTGCTGGACGAGCCGACCAACCAATTGGATAATGACGCGGATTA
TTTGGCTGGAATACTGCTCAAGCGTTTGAAGGCAGCTTGGTCTGTGATTACCCACGACCGCTGTTTTTTGGATAATATCGC
CACGCGAATTGTCGAACCTTGACC
abcZ WGS: automated extract (BIGSdb)
TTTGATACCGTTGGCGAAGSTTTGGSCGAAATTCGCGATTITATTCGCCGCTTATCATGTGAGCCATGAGTTGGAGCACA
ATTGACAGCAGCTTTTATTAAGAACTCAACGAATTACAACTCGAAATCGAAGCGAAGGACGGCTGGAACTGGATGCGGC
AGTCAAGCAGACTTTGGGCGAACTCGGTTTGGCGAAAAATCGGCAACCTTTCCGGCGGTGAGAAAAAGCGGTGTC
GCTTGGCGCAGGCTTTGGGTGCAAAAGCCCGACGTAATTGCTGCTGGACGAGCCGACCAACCAATTGGATAATGACGCGGATTA
TCTGGTTGGAACCTGCTCAAGCGTTTGAAGGCAGCTTGGTCTGTGATTACCCACGACCGCTTTTTTTGGATAATATCGC
CACGCGAATTGTCGAACCTTGACC
```

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

PubMLST Database home Curator home Contents

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

## Batch insert sequences

**Import status**

Sender: Keith Jolley

No obvious problems identified so far.

Action

**Import data**

**Data to be imported**

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

locus	allele_id	sequence	status	type_allele	sender	curator	date_entered	timestamp	comments	flags
abcZ	878	TTTGATACTGTTGCCGAAGG ... GCGAATTGTCGAACTTGACC	WGS: automated extract (BIGSdb)		2	2	2018-06-07	2018-06-07		
abcZ	879	TTTGATACCGTTGCCGAAGG ... GCGGATTGTCGAACTTGACC	WGS: automated extract (BIGSdb)		2	2	2018-06-07	2018-06-07		

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

PubMLST Database home Curator home Contents

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

## Batch insert sequences

**Import status**

Primary key	Problem(s)
locus: abcZ; allele_id: 878	Sequence contains non nucleotide (A C G T) characters.

**Data to be imported**

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

locus	allele_id	sequence	status	type_allele	sender	curator	date_entered	timestamp	comments	flags
abcZ	878	TTTGATACTGTTGCCGAAGG ... GCGAATTGTCGAACTTGACC	WGS: automated extract (BIGSdb)		2	2	2018-06-07	2018-06-07		
abcZ	879	TTTGATACCGTTGCCGAAGG ... GCGGATTGTCGAACTTGACC	WGS: automated extract (BIGSdb)		2	2	2018-06-07	2018-06-07		

### 6.2.2.2 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.

PubMLST Database home Curator home Contents

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

## Database curator's interface - Neisseria profile/sequence definitions

Curator functions

Users Sequences MLST profiles

FASTA

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

PubMLST
Database home
Curator home
Contents

Logged in: **Keith Jolley (keith)**
Log out
Change password
Help
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
TTTGATACGTGGCCGAAGGTTTGGGCGAAATTCGCGATTATTGCGCCGTTATCATCAT
GTCAGCCATGAGTTGGAAAATGGTTCGAGTGAGGCTTTGTTGAAAGAACTCAACGAATTG
CAACTTGCAATCGAAGCGAAGGACGGCTGGAACTGGATGCGGCAGTCAAGCAGACTTTG
GGGAACTCGGTTTGGCGAAAATGAAAAATCGGCAACCTTTCGCGCGTCAAGAAAAG
CGCGTCGCCCTTGGCGCAGGCTTGGGTGCAAAAGCCCGACGTATTGCTGCTGGACGAGCCG
ACCAACCAITTTGGATATCGACGCGATTATTGGCTGGAAAATCTGCTCAAAGCGTTTGA
GGCAGCTTGGTCGTGATTACCCACGACCGTCGTTTTTGGACAATATCGCCACGCGAATT
GTCGAACTTGACC
>isolate2
TTTGATACGTTGGCCGAAGGTTTGGGCGAAATTCGCGATTATTGCGCCGTTATCATCAT

```

☐ 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'.

PubMLST Database home Curator home Contents

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

## Batch insert sequences

**Sequence check**

Locus: abcZ

Original designation	Allele id	Status	Action
isolate1	878	OK	<a href="#">Upload valid sequences</a>
isolate2	879	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.

PubMLST Database home Curator home Contents

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

## Batch insert sequences

**Sequence check**

Locus: abcZ

Original designation	Allele id	Status	Action
isolate1	878	OK	<a href="#">Upload valid sequences</a>
isolate2	879	Sequence contains non nucleotide (A C G T) 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 update/delete sequences link.



Either search for specific attributes in the search form, or leave it blank and click 'Submit' to return all alleles. For a specific allele, select the locus in the filter (click the small arrow next to 'Filter query by' to expand the filter) and enter the allele number in the allele\_id field.

The screenshot shows the 'Query sequences for Neisseria profile/sequence definitions database'. It includes a header with navigation links and a status bar. The main content area has a title 'Query sequences for Neisseria profile/sequence definitions database'. Below the title, there's a paragraph explaining that some loci have additional fields not searchable from this page. Another paragraph notes that some loci have allele ids defined as text strings. A section titled 'Please enter your search criteria below (or leave blank and submit to return all records)' contains a search form. The 'Search criteria' section has a dropdown for 'allele id' set to '=', a text input for '4', and a '+' button. The 'Display' section has 'Order by' set to 'locus' and 'ascending', and 'Display' set to '25 records per page'. The 'Filter query by' section is expanded, showing fields for 'locus' (set to 'abcZ'), 'status', 'type allele', 'sender', 'curator', and 'allele flag'. There are 'Reset' and 'Submit' buttons in the 'Action' section.

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

PubMLST Database home Curator home Contents

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

### Query sequences for Neisseria profile/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:  
 type allele:  
 sender:  
 curator:  
 allele flag:

**Action**  
 Reset Submit

1 record returned.

Delete Database configuration Flags  
 Delete ALL Export configuration/data Batch set

Delete	Update	locus	allele id	sequence	sequence length	status	type allele	sender	curator	date entered	timestamp	comments	flags
<input checked="" type="checkbox"/>	<input type="checkbox"/>	abcZ	4	TTTGATACCGTTGCC ... TTGTCGAACTCGATC	433	Sanger trace checked	<input type="checkbox"/>	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. You can also choose to delete and retire the allele identifier. If you do this, the allele identifier will not be re-used.

PubMLST Database home Curator home Contents

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

### Delete allele sequence

You have chosen to delete the following record. Select 'Delete and Retire' to prevent the identifier being reused.

locus: abcZ  
 allele id: 4  
 sequence: TTTGATACCG TTGCGAAGG TTGGGCGAA ATTCTGTATT TATTGCGCC TTATCATCAT GTCAGCCATG AGTTGGAAAA TGGTTCGAGT GAGGCTTTGT  
 TGAAGAAGCT CAACGAATTG CAACTTGAAA TCGAAGCGAA GGACGGCTGG AACTTGGATG CGGCAGTCAA GCAGACTTTG GGGGAACTCG GTTTGCCGGA  
 AAATGAAAA ATCGGCAACC TTTCGGCGG TCAGAAAAAG CGCGTCGCCT TGCTCAGGC TTGGGTGCAA AAGCCCGACG TATTGCTGCT GGACGAGCCG  
 ACCAACCATT TGGATATCGA CGCGATTATT TGGCTGGAAA ATCTGCTCAA AGCGTTTGAA GGCAGCTTGG TTGTGATTAC CCACGACCCG CGTTTTTTTG  
 ACAATATCGC CACGCGGATT GTCGAACTCG ATC  
 status: Sanger trace checked  
 type allele: false  
 sender: Keith Jolley  
 curator: Man-Suen Chan  
 date entered: 2001-02-07  
 timestamp: 2009-11-11  
 comments:

**Action**  
 Delete Delete and Retire

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

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

PubMLST
Database home
Curator home
Contents

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
ATTCTGTGATT TATTGCGCCG TTATCATCAI
GTCAGCCATG AGTTGGAAAA TGGTTCGAGT
GAGGCTTTGT TGAAGAACT CAACGAATTG
CAACTTGAAA TCGAAGCGAA GGACGGCTGG
AAACTGGATG CGGCAGTCAA GCAGACTTTG
GGGGAACTCG GTTTCGCCGA AAATGAAAA
```

  
status: Sanger trace checked  
sender: Jolley, Keith (keith)  
curator: **Keith Jolley (keith)**  
date entered: 2001-02-07  
datestamp: 2018-06-07  
type allele: ☐ true ☐ false
New allele searches can be constrained to use just type alleles in comparisons  
comments:
  
Flags:

- atypical
- contains IS element
- downstream fusion
- frameshift
- internal stop codon

Use Ctrl click to select/deselect multiple choices  
PubMed ids:
  
ENA ids:
  
Genbank ids:
  
☐ Override sequence length check

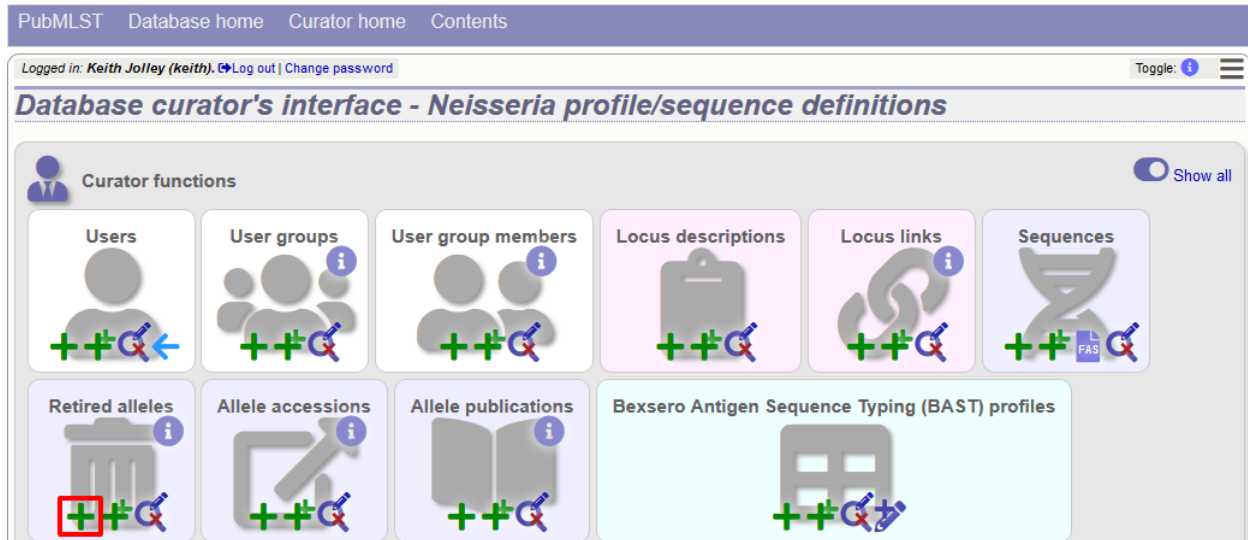
Action
Reset
Submit

## 6.4 Retiring allele identifiers

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

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





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

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

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

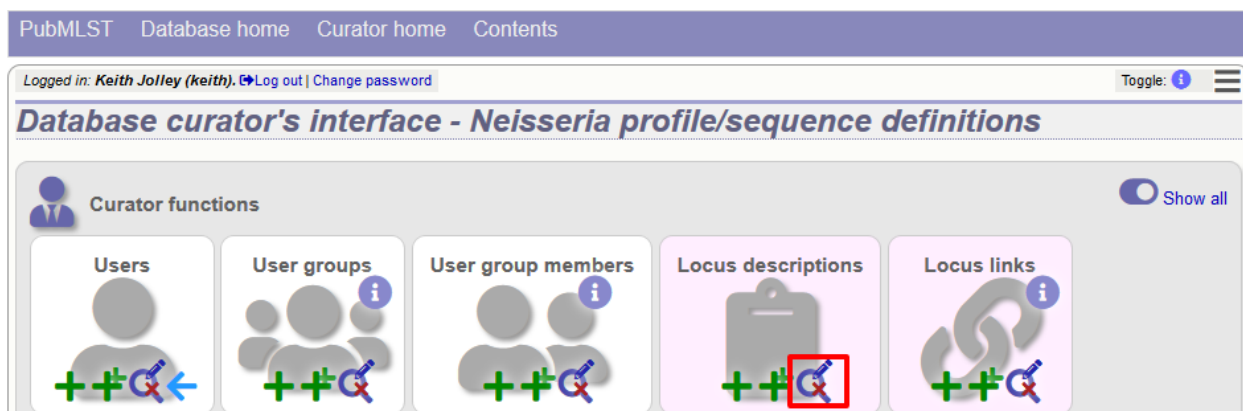
## 6.5 Updating locus descriptions

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

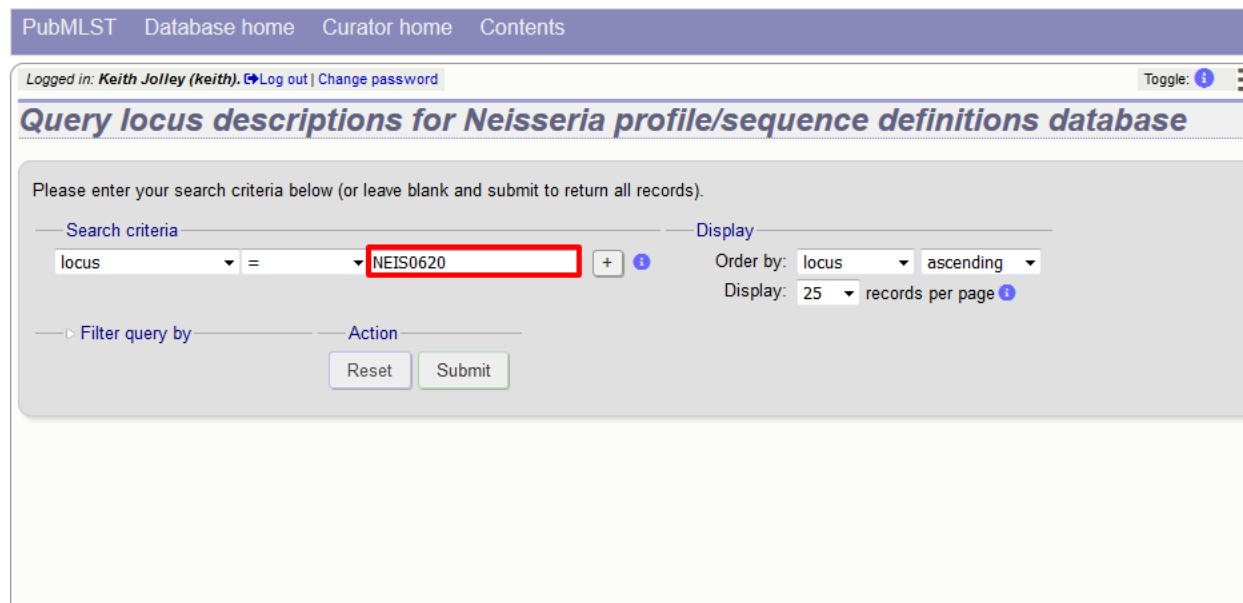
**Note:** In recent versions of BIGSdb, a blank description record is created when a new locus is defined. The following

instructions assume that this is the case. It is possible for this record to be deleted or it may never have existed if the locus was created using an old version of BIGSdb. If the record does not exist, it can be added by clicking the Add (+) button in the 'locus descriptions' box. Fill in the fields in the same way as described below.

To edit a locus description, first you need to find it. Click the update/delete button in the 'locus descriptions' box on the sequence database curator's page (depending on the permissions set for your user account not all the links shown here may be displayed). This function is normally hidden, so you may need to click the 'Show all' toggle to display it.




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



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


PubMLST Database home Curator home Contents

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

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


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

**Search criteria**


locus:  =  + 


**Display**


Order by: locus ascending

Display: 25 records per page 

**Filter query by**

locus:  

curator:  

common name:  


**Action**

Reset Submit

Click 'Submit'.

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


PubMLST Database home Curator home Contents

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

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


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

**Search criteria**


locus:  =  + 


**Display**


Order by: locus ascending

Display: 25 records per page 

**Filter query by**

locus:  

curator:  

common name:  



**Action**

Reset Submit

1 record returned.

**Delete** **Database configuration**

Delete ALL Export configuration/data

Delete	Update	locus	full name	product	description	curator	timestamp
		NEIS0620		malate oxidoreductase (EC 1.1.1.38)	Final step in TCA cycle producing oxaloacetate.	Keith Jolley	2010-10-28

Fill in the form as needed:

PubMLST
Database home
Curator home
Contents

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

locus: **NEIS0620**  
curator: **Keith Jolley (keith)**  
timestamp: **2018-06-07**  
full name:   
product: **malate oxidoreductase (EC 1.1.1.38)**  
description: **Final step in TCA cycle producing oxaloacetate.**  
aliases:   
  
  
PubMed ids:   
links:   
(Format: URL|description)

Action

Reset Submit

- 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

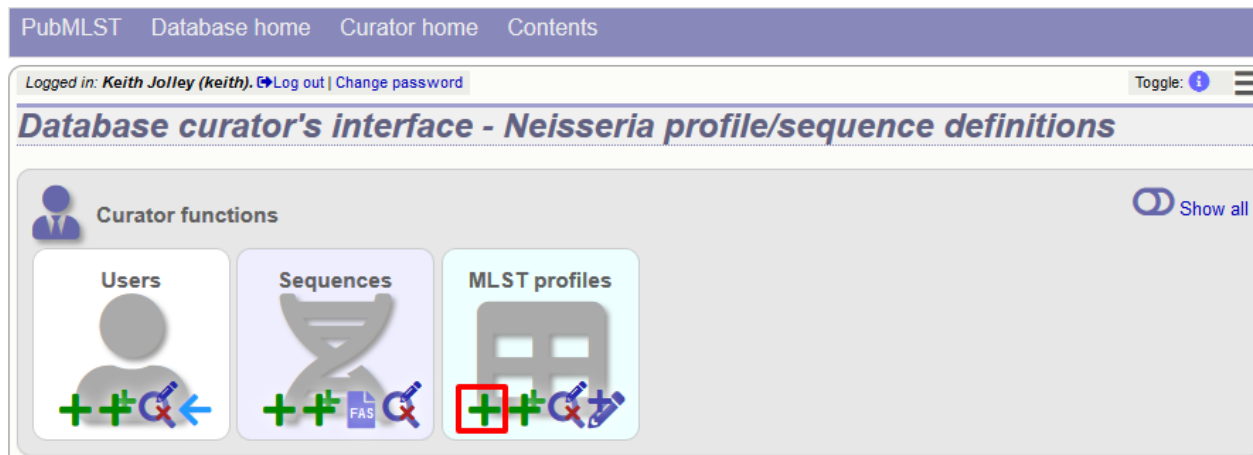
Enter links to additional web-based resources. Enter the URL first followed by a pipe symbol (|) and then the description.

Click 'Submit' when finished.

## 6.6 Adding new scheme profile definitions

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

To add a single profile you can click the add (+) profiles link in the box named after the scheme name (e.g. MLST):



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

PubMLST Database home Curator home Contents

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: ! 3015

abcZ: ! 2

adk: ! 3

aroE: ! 4

fumC: ! 122

gdh: ! 8

pdhC: ! 4

pgm: ! 6

sender: ! Jolley, Keith (keith)

clonal\_complex:

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

date\_entered: ! 2018-06-07

datestamp: ! 2018-06-07

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:

PubMLST Database home Curator home Contents

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

## Database curator's interface - *Neisseria* profile/sequence definitions

**Curator functions** [Show all](#)

Users

Sequences

MLST profiles

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

PubMLST Database home Curator home Contents

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)

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'.

PubMLST Database home Curator home Contents

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)

```

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 previously defined profiles

☐ Ignore duplicate profiles

Action

Reset Submit

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

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith) Log out Change password Help Toggle

## Batch insert MLST profiles

### Import status

Sender: Keith Jolley

No obvious problems identified so far.

Action

**Import data**

### Data to be imported

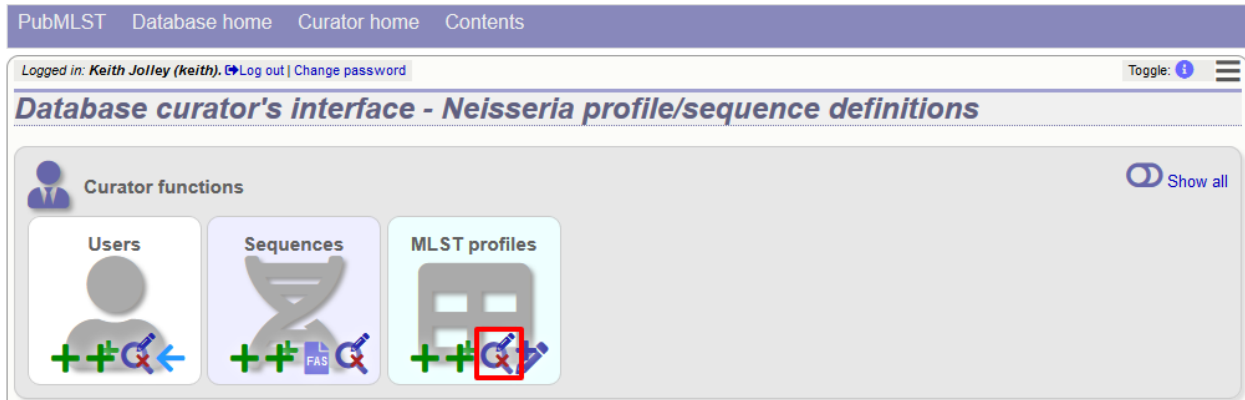
The following table shows your data. Any field coloured red has a problem and needs to be checked.

ST	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	clonal_complex	sender	curator	date_entered	timestamp
3015	2	3	4	122	8	4	6		2	2	2018-06-07	2018-06-07

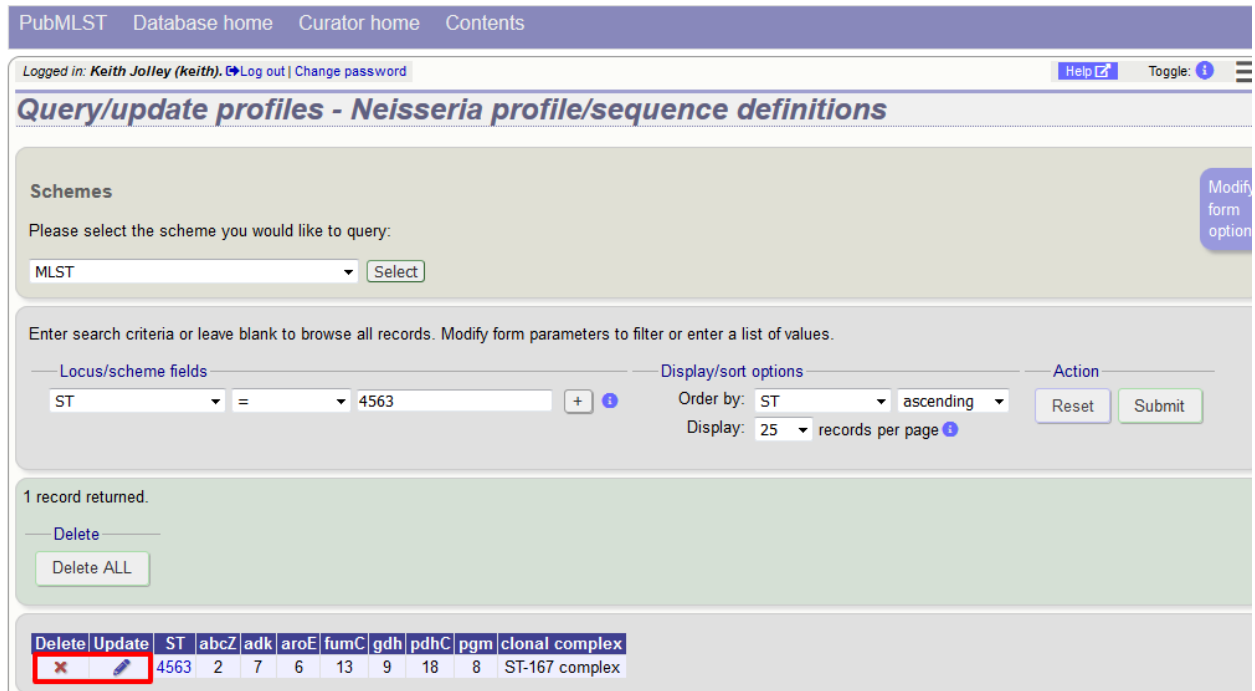


## 6.7 Updating and deleting scheme profile definitions

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



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



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

For deletion of a single record, the full record will be displayed. Confirm deletion by clicking 'Delete'. You can also choose to delete and retire the profile identifier. If you do this, the profile identifier will not be re-used.

PubMLST Database home Curator home Contents

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

## Delete profile

You have chosen to delete the following record. Select 'Delete and Retire' to prevent the identifier being reused.

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

Action

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

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith) Log out Change password Help Toggle

## Update profile

Record

Update your record as required - required fields are marked with an exclamation mark (!):

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: ! 2018-06-07  
 PubMed ids:

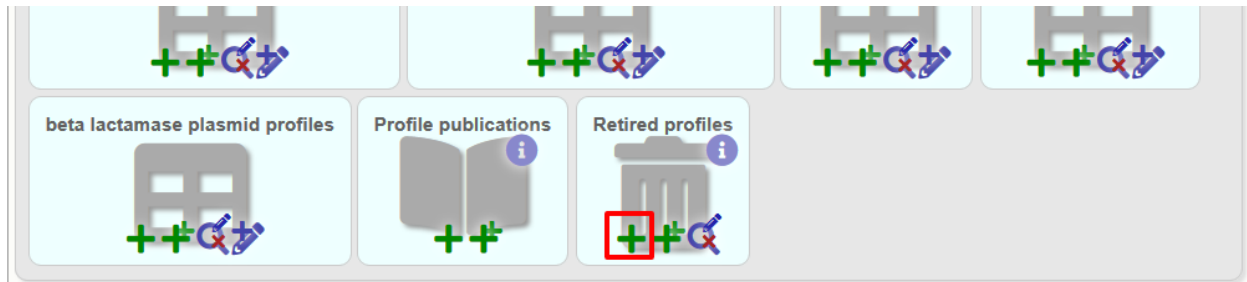
Action

## 6.8 Retiring scheme profile definitions

Sometimes there is a requirement to prevent the automated assignment of a particular profile identifier (e.g. ST) - a profile with that identifier may have been commonly used and has since been removed. Reassignment of the identifier

to a new profile may lead to confusion, so in this instance, it would be better to prevent this.

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



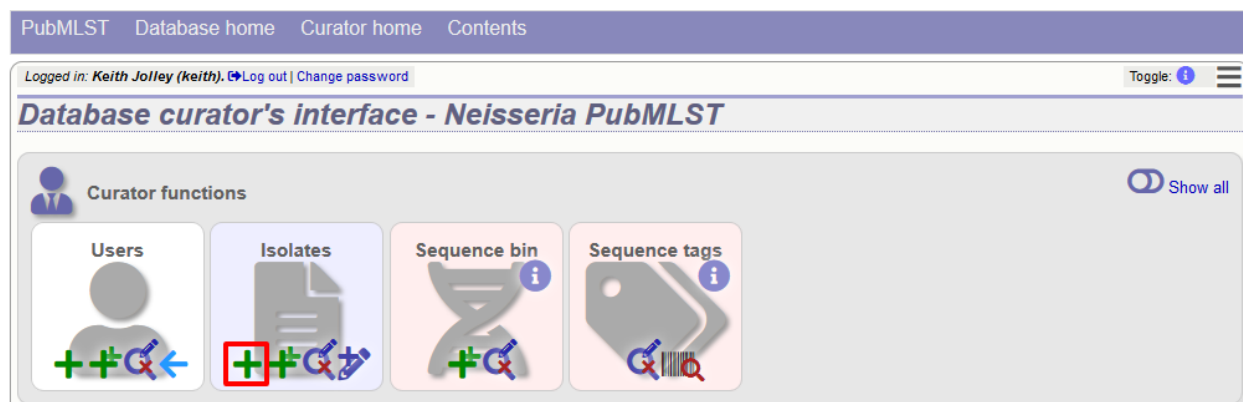
Select the scheme from the dropdown list box and enter the profile id. Click ‘Submit’.

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

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

## 6.9 Adding isolate records

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



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

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

**Isolate fields**

id: 60465 (!)

isolate: J323\_2 (!)

country: UK (!)

species: Neisseria meningitidis (!)

sender: Jolley, Keith (keith) (!)

curator: Keith Jolley (keith) (!)

date entered: 2018-06-07 (!)

datestamp: 2018-06-07 (!)

region: (!)

year: 2014 (!)

epidemiological year: (!)

age yr: (!)

age mth: (!)

sex: (!)

disease: meningitis (!)

source: CSF (!)

epidemiology: (!)

serogroup: B (!)

genogroup: (!)

MLEE designation: (!)

serotype: (!)

sero subtype: (!)

ET no: (!)

penicillin: (!)

penicillin range: (!)

**Allele designations**

**MLST**

abcZ	adk	aroE	fumC
gdh	pdhC	pgm	

**Finotyping antigens**

PorA VR1	PorA VR2	PorA VR3	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.



Download a submission template in Excel format from the link.

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).

PubMLST Database home Curator home Contents

Logged in: **Keith Jolley (keith)**. [Log out](#) | [Change password](#) Toggle: 1

## 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:

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). Action

Reset Submit

```
isolate aliases references country region year epidemiological_year
age_yr age_mth sex disease source epidemiology species serogroup
genogroup 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 aroE fumC
gdh pdhC pgm PorA_VR1 PorA_VR2 FetA_VR gyrA penA
rpoB
J323_2 UK 2014
meningitis CSF Neisseria meningitidis
B
```

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

PubMLST Database home Curator home Contents

Logged in: **Keith Jolley (keith)**. [Log out](#) | [Change password](#) Toggle: 1

## Batch insert isolates

### Import status

Sender: Keith Jolley

No obvious problems identified so far.

Action

**Import data**

### Data to be imported

The following table shows your data. Any field with red text has a problem and needs to be checked.

id	isolate	aliases	references	country	region	year	epidemiological_year	age_yr	age_mth	sex	disease	source	epidemiology	species	serogroup	ge
60465	J323_2			UK		2014					meningitis	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 home Curator home Contents

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

## Batch insert isolates

**Import status**

Primary key	Problem(s)
id: 60465	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	gen
60465	J323_2			UK		2014					meningitis	CSF		Neisseria meningitidis	B	

## 6.10 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 home Curator home Contents

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

## Database curator's interface - Neisseria PubMLST

Curator functions Show all

Users

Isolates

Sequence bin

Sequence tags

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

PubMLST Database home Curator home Contents

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

## Isolate query/update

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields

Combine with: **AND**

country = USA +

year = 2009

Display/sort options

Order by: **id** ascending

Display: **25** records per page

Action

[Reset](#) [Submit](#)

30 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Delete Tag scanning Projects

[Delete ALL](#) [Scan](#) [Select project...](#) [Link](#)

Page: [1](#) [2](#) [>](#) [Last](#)

Delete	Update	Sequence bin	New version	Isolate fields										Seqbin size (bp)	Conti
				id	isolate	aliases	country	year	disease	species	serogroup	genogroup	capsule group		
<a href="#">✕</a>	<a href="#">✎</a>	<a href="#">📁</a>	<a href="#">+</a>	12674	M18700		USA	2009		Neisseria meningitidis	B		B	0	0
<a href="#">✕</a>	<a href="#">✎</a>	<a href="#">📁</a>	<a href="#">+</a>	12675	M18701		USA	2009		Neisseria meningitidis	B		B	0	0
<a href="#">✕</a>	<a href="#">✎</a>	<a href="#">📁</a>	<a href="#">+</a>	12676	M18725		USA	2009		Neisseria meningitidis	B		B	0	0
<a href="#">✕</a>	<a href="#">✎</a>	<a href="#">📁</a>	<a href="#">+</a>	13090	M19024	PA09015	USA	2009	meningitis	Neisseria meningitidis	B		B	0	0
<a href="#">✕</a>	<a href="#">✎</a>	<a href="#">📁</a>	<a href="#">+</a>	14627	M20918		USA	2009		Neisseria meningitidis	A		A	1717803	2046
<a href="#">✕</a>	<a href="#">✎</a>	<a href="#">📁</a>	<a href="#">+</a>	14998	M21319		USA	2009	invasive (unspecified/other)	Neisseria meningitidis	B		B	0	0
<a href="#">✕</a>	<a href="#">✎</a>	<a href="#">📁</a>	<a href="#">+</a>	15000	M21323		USA	2009	invasive	Neisseria	B		B	0	0

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




PubMLST Database home Curator home Contents


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

## Delete isolate


You have chosen to delete the following record. Select 'Delete and Retire' to prevent the isolate id being reused.

 **Provenance/meta data**

id: 14627	region: IA	sender: Xin Zhao, Novartis (formerly at US CDC)
isolate: M20918	year: 2009	curator: Auto Tagger
strain designation: A: P1.20.9: F3-1: ST-4789,7980 (cc5)	source: CSF	update history: <a href="#">83 updates</a> <a href="#">show details</a>
country: USA	species: Neisseria meningitidis	date entered: 2010-01-26
continent: North America	serogroup: A	datestamp: 2018-01-31
	capsule group: A	

 **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: 2,046	N50 contig number: 510	N95 contig number: 1,633
total length: 1,717,803 bp	N50 length (L50): 1,155	N95 length (L95): 324
max length: 5,436 bp	N90 contig number: 1,414	loci tagged: 1,518
mean length: 840 bp	N90 length (L90): 456	detailed breakdown: <a href="#">Display</a>

**Action**

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'.

PubMLST Database home Curator home Contents

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: 2018-06-07

region: IA

year: 2009

epidemiological year:

age yr:

age mth:

sex:

disease:

source: CSF

epidemiology:

serogroup: A

genogroup:

MLEE designation:

serotype:

sero subtype:

ET no:

penicillin:

penicillin range:

amoxicillin:

Loci

- All loci
  - Capsule
  - Genetic Information Processing
  - Genomic islands
  - Lineage Schemes
  - Metabolism
  - N. gonorrhoeae AMR
  - 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.

PubMLST Database home Curator home Contents

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

date stamp: 2018-06-07

region: IA

year: 2009

epidemiological year:

age yr:

age mth:

sex:

disease:

source: CSF

epidemiology:

serogroup: A

genogroup:

MLEE designation:

serotype:

sero subtype:

ET no:

penicillin:

penicillin range:

amoxicillin:

Loci

- All loci
  - Capsule
  - Genetic Information Processing
  - Genomic islands
  - Lineage Schemes
  - Metabolism
  - N. gonorrhoeae AMR
  - Pilin
  - Typing
  - Other schemes
  - Loci not in schemes

Navigate and select schemes within tree to display allele designations

Locus: '16S\_rDNA

Add/update

PubMLST Database home Curator home Contents

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

date stamp: 2018-06-07

region: IA

year: 2009

epidemiological year:

age yr:

age mth:

sex:

disease:

source: CSF

epidemiology:

serogroup: A

genogroup:

MLEE designation:

serotype:

sero subtype:

ET no:

penicillin:

penicillin range:

amoxicillin:

sulphonamide:

ceftriaxone:

ceftriaxone range:

chloramphenicol:

chloramphenicol range:

Loci

- All loci
  - Capsule
  - Genetic Information Processing
  - Genomic islands
  - Lineage Schemes
  - Metabolism
  - N. gonorrhoeae AMR
  - Pilin
  - Typing
    - MLST
    - Finotyping antigens
    - Antigen genes
    - N. gonorrhoeae cgMLST v1.0
    - N. meningitidis cgMLST v1.0
    - Ribosomal MLST
    - eMLST (20 locus partial genes)
    - eMLST (20 locus whole genes)
    - rplF species
    - Other schemes
  - Loci not in schemes

MLST

abc7	adk	aroE	fumC	gdh	pdhC	pgm	ST	clonal complex
1	1	2	1	3	6	334	19	4789, 7980

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 home Curator home Contents

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

### 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: 2018-06-07

region: IA

year: 2009

epidemiological year:

age yr:

age mth:

sex:

disease:

source: CSF

epidemiology:

serogroup: A

genogroup:

MLEE designation:

serotype:

sero subtype:

**Loci**

All loci

- Capsule
- Genetic Information Processing
- Genomic islands
- Lineage Schemes
- Metabolism
- N. gonorrhoeae AMR
- 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 home Curator home Contents

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

capsule group: A

sender: Xin Zhao

curator: Auto Tagger

update history: 63 updates show details

date entered: 2010-01-26

datestamp: 2018-01-31

**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: 2018-06-07

date entered: 2018-06-07

comments:

**Action**

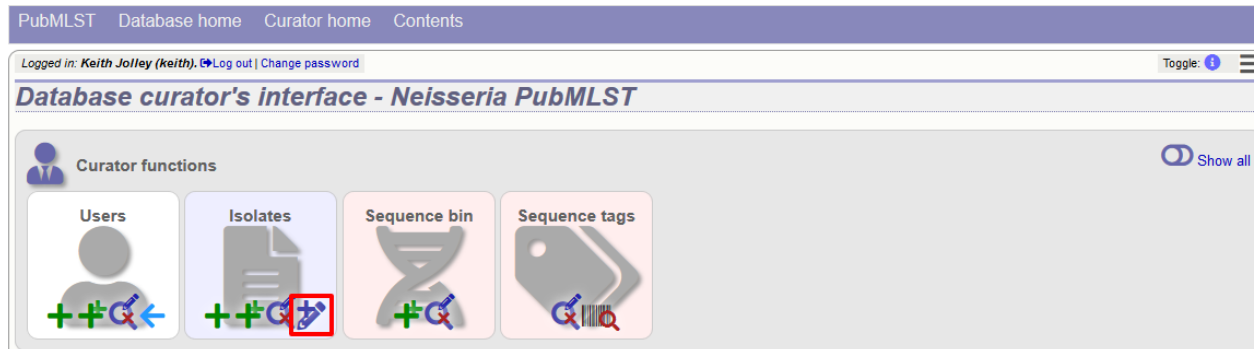
Reset Submit

**Existing designations**

Update	Delete	allele id	sender	status	method	comments
		1	Xin Zhao	confirmed	manual	

## 6.11 Batch updating multiple isolate records

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



Prepare your update data in 3 columns in a spreadsheet:

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

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

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

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

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

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

PubMLST Database home Curator home Contents

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

## 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/locus 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.

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 home Curator home Contents


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

## 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



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

## 6.12 Deleting multiple isolate records

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

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

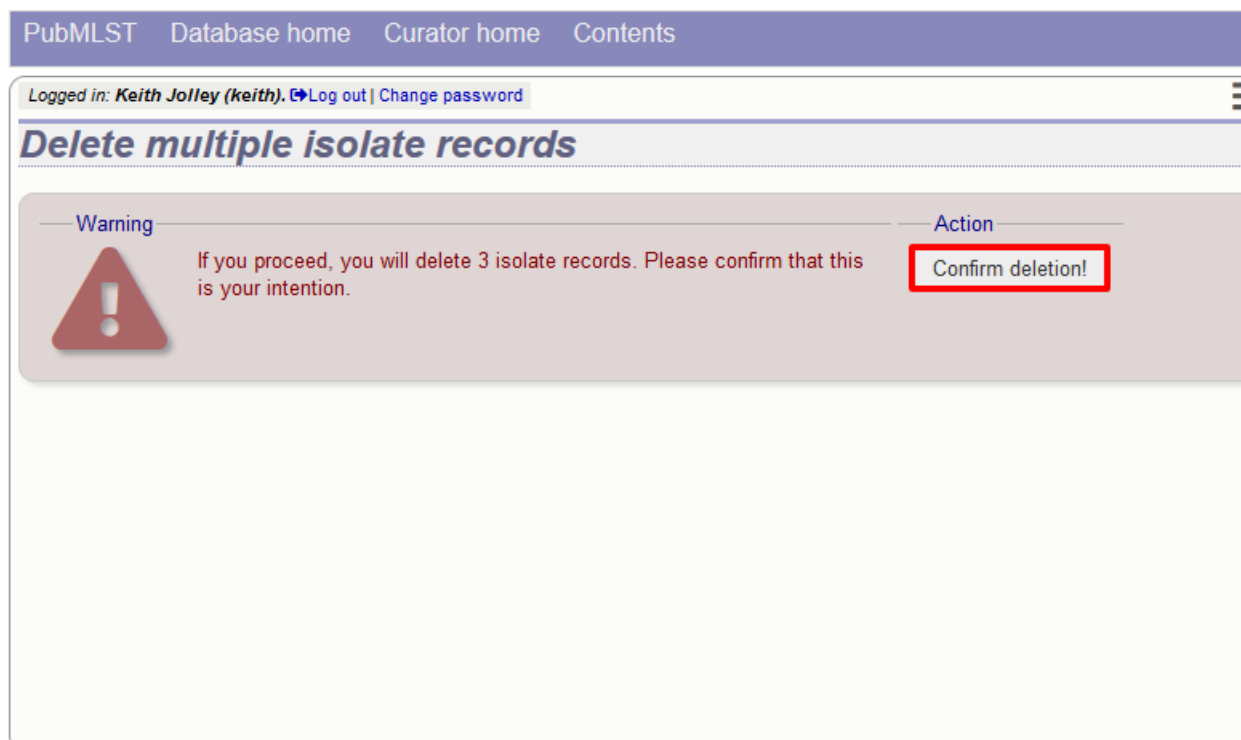


Enter search criteria that specifically return the isolates you wish to delete. Click 'Delete ALL'.

The screenshot shows the 'Isolate query/update' page. It includes search criteria for 'date entered' (2014-03-18) and 'curator (surname)' (Jolley). The 'Action' section has 'Reset' and 'Submit' buttons. Below the search form, it states '3 records returned'. Under the 'Delete' tab, the 'Delete ALL' button is highlighted with a red rectangular box. Below this is a table of 3 records.

Delete	Update	Sequence bin	New version	id	isolate	aliases	country	year	disease	species	serogroup	genogroup	capsule group	PorA VR1	PorA VR2	FetA VR
				28787	M22296		USA			Neisseria meningitidis				5-1	2-81	
				28788	M22553		USA			Neisseria meningitidis				5-2	10-96	
				28789	M22568		USA			Neisseria meningitidis				7-2	4-39	

You will have a final chance to change your mind:

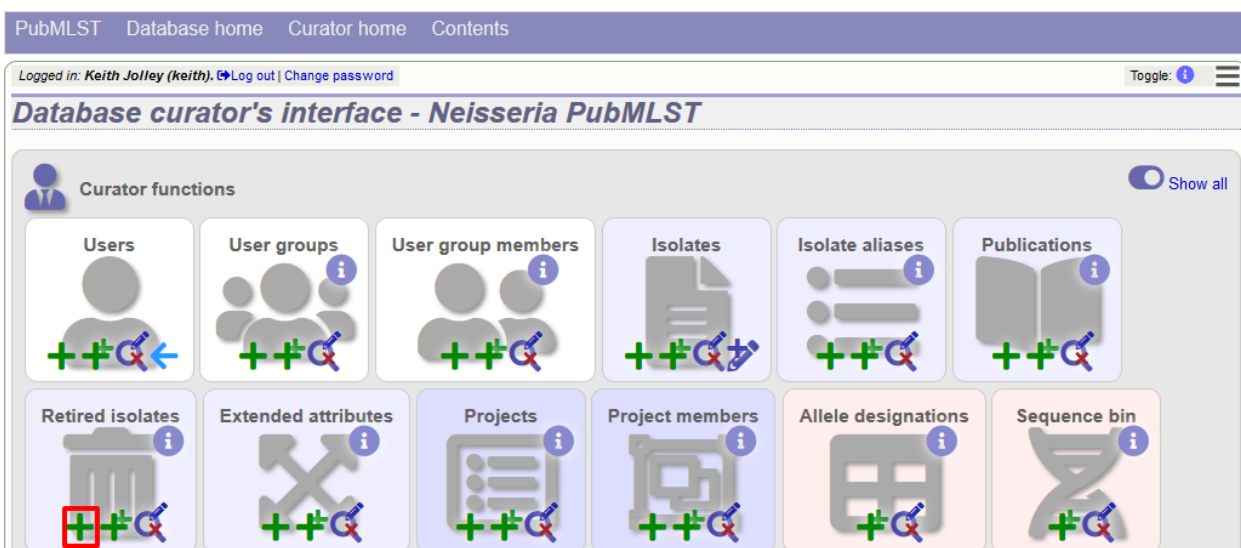


Click 'Confirm deletion!'.

## 6.13 Retiring isolate identifiers

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

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





Enter the isolate id to retire and click ‘Submit’.

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

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

## 6.14 Setting alternative names for isolates (aliases)

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

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

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

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

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

PubMLST Database home Curator home Contents

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

### Batch insert isolate aliases

This page allows you to upload isolate alias 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).

```
isolate_id    alias
5473         JHS212
5473         MN11
```

Action

Reset Submit

Back

A confirmation page will be displayed.

PubMLST Database home Curator home Contents

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

### Batch insert isolate aliases

**Import status**

No obvious problems identified so far.

Action

Import data

**Data to be imported**

The following table shows your data. Any field with red text has a problem and needs to be checked.

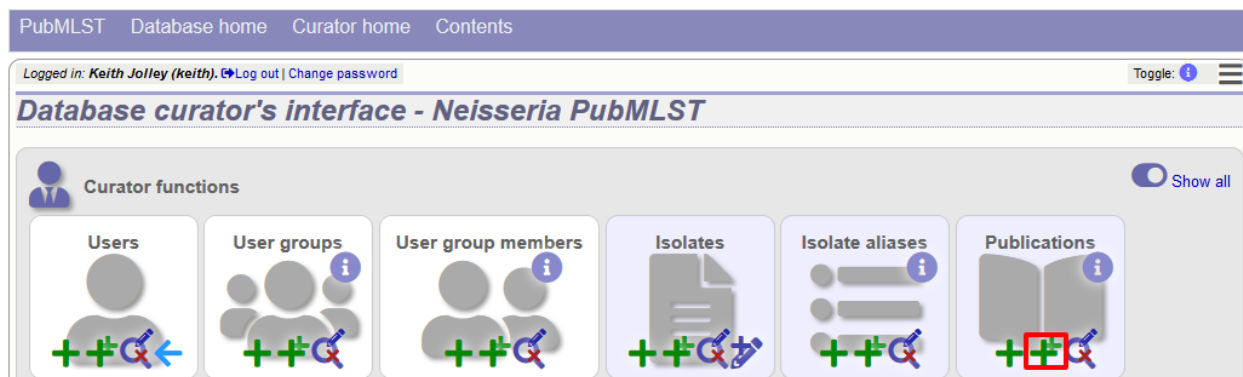
isolate_id	alias	timestamp	curator
5473	JHS212	2016-12-16	2
5473	MN11	2016-12-16	2

Click 'import data'.

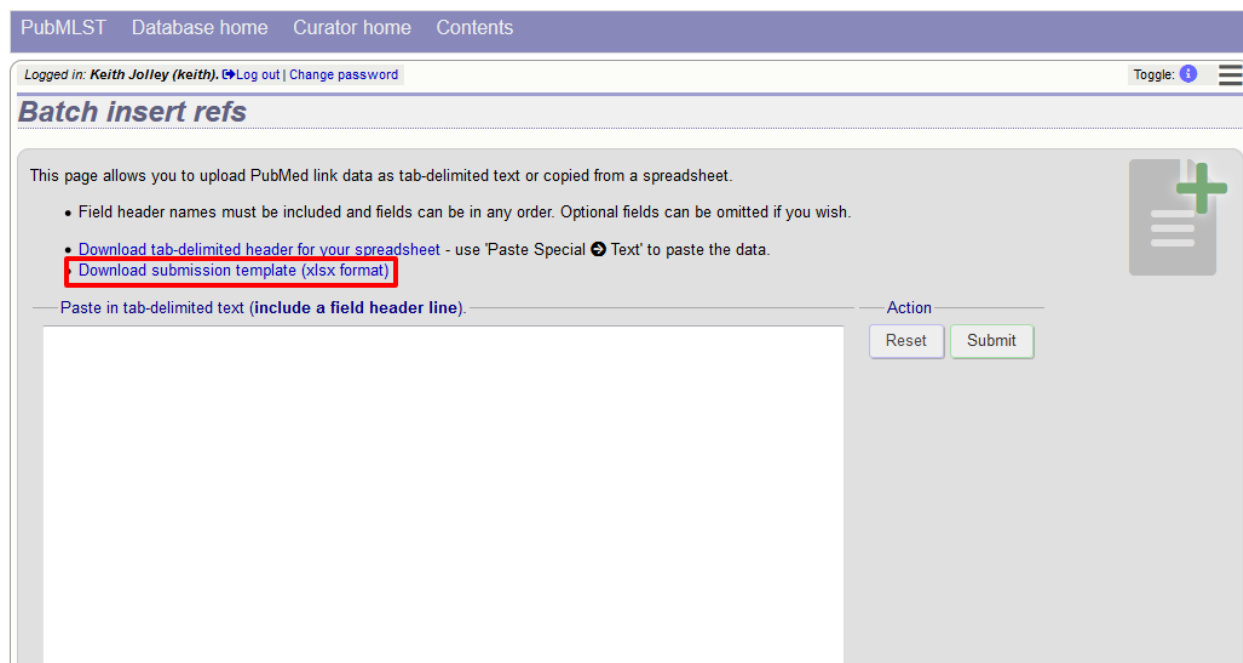
## 6.15 Linking isolate records to publications

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

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



Open the Excel template by clicking the link.



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

PubMLST Database home Curator home Contents

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

### Batch insert refs

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

- Field header names must be included and fields can be in any order. Optional fields can be omitted if you wish.
- 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).

isolate_id	pubmed_id
6160	17517841
6162	17517841
18968	26515523

Action

Reset Submit

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

## 6.16 Uploading sequence contigs linked to an isolate record

### 6.16.1 Select isolate from drop-down list

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

PubMLST Database home Curator home Contents

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

### Database curator's interface - Neisseria PubMLST

Curator functions

Users Isolates Sequence bin Sequence tags

Select the isolate that you wish to link the sequence to from the dropdown list box (or if the database is large and there are too many isolates to list, enter the id in the text box). You also need to enter the person who sent the data. Optionally, you can add the sequencing method used.

Paste sequence contigs in FASTA format in to the form.

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith). Log out | Change password Toggle: 1

## Upload 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:

```

ACCGTCAAAATCGTCGGCATCGATGAATCGATACCGCGCAAAACAAAATTTCTGGATT
TCGCCGCTGGCTGGTTGTCGATTAAAGCGCGTGAAGGCGGACGAAGTGGTTTAAACACG
CCGGAAGGCGGTGAGGAAATGAAATCCTTTCTGCGAATACATCAAAATCGACATGATCG
GGAATGCGCTCTGAAGCGGACAAACGCGCGCTTTCCCAATAAGGAATACGAATGAG
GAAATCCGCGCAAAATCTCGGGATCGACACACGGAAGCGCACTGTATGCGGCCCA
CGCGGCGCAGACSCATTGGACTGGTTTTTACCCGCCAAAGCCCCGCGCTGTCGACAT
CATTAAAGCACAATAATCACCGCGCATCGCGCGCTTTGTCAGCGTTGTCGCCCTTTT
CGTCAAGGAAGCGCGCAAAACATCGCGCGCATCTTCCGGAATACCGATACACATCAT
CCAATTCACGCGGACGAAGACGACGATCTGCGCGCAGTTCCACCGCCCTATATCAA
GGCCATTCGTTGTCAGCGGATCAGACATCGAAACGCCCGCGCGCTTCCCGACGC
TCAGGCATGCTGTTGATGCTTACCATCTCTCGGAATACGGCGGACCGGACACGCTT
CGACTGGACGCTGTTGCGGAATATTCGGGCAACCGTGGGTGCTTCCCGCGGCGTGAC
CCTGAAACGTCGACGAAGCCATCGCATCACCGGAGCGGAAGCGGTGATGTATCGCG
CGCGTGGGAAGCGTCTAAAGCAAAAGACCGGCAAAATTTGCCGCTTTATCGCAAC
CGCCACCGCCTATCCCGTTAAAGCACAAAATTTGCCCGGAATGACTTATAGTGGAT
TAAGAAACCGCTCGCGCTGCTGCTAGCTCAAGGGAAGCATCTCTAAGGTG
CTGAAGCACCAAGTAATCGGTCGCTACTGCTGCTGCTGCGGCTTCTGCGGCTTTC
TCCTGATTTTGTAAATCCACTATAATCTAAATAATATGCTATTAATCAGTAATTC
TGATGAATTTGAAACCTTAATCCCGCTATTCCGCTCAGCGGGAATCCGTTCAATGA
GTTTCAGTATTAGATAAATTTTGAATCTGA

```

Attributes

isolate id: 2

sender: Jolley, Keith (keith)

method:

run id:

assembly id:

Options

☐ Don't insert sequences shorter than 25 bps.

Link to experiment:

Alternatively upload FASTA file or enter Genbank accession

Select FASTA file:

Browse... No file selected.

Action

Reset Submit

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

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (keith). Log out | Change password Toggle: 1

## Upload sequences

The following sequences will be entered.

Original designation	Sequence length	Comments
180062 NODE_116_length_5370_cov_18.338547	5408	
180063 NODE_267_length_64_cov_54.562500	102	
180064 NODE_1024_length_456_cov_16.434210	494	
180065 NODE_367_length_3545_cov_22.858955	3583	
180066 NODE_361_length_87_cov_17.862068	125	
180067 NODE_1617_length_297_cov_11.111111	335	
180068 NODE_909_length_95_cov_9.073684	133	
180069 NODE_699_length_98_cov_110.918365	136	
180070 NODE_553_length_84_cov_47.964287	122	
180071 NODE_182_length_6772_cov_19.882162	6810	
180072 NODE_928_length_347_cov_35.201729	385	
180073 NODE_19_length_12542_cov_19.259449	12580	
180074 NODE_60_length_5125_cov_18.960781	5163	
180075 NODE_168_length_7439_cov_17.615808	7477	
180076 NODE_1041_length_109_cov_27.752293	147	
180077 NODE_71_length_2120_cov_19.594339	2158	
180078 NODE_318_length_1827_cov_15.821566	1865	
180079 NODE_207_length_6398_cov_19.577681	6436	
180080 NODE_162_length_13775_cov_18.618221	13813	
180081 NODE_664_length_160_cov_35.068748	198	
180082 NODE_56_length_9475_cov_20.244328	9513	
180083 NODE_356_length_297_cov_41.383839	335	
180084 NODE_778_length_582_cov_42.140892	620	
180085 NODE_137_length_10168_cov_16.055567	10206	
180086 NODE_1102_length_82_cov_43.329269	120	
180087 NODE_227_length_111111_cov_11.111111	111111	

Summary

Number of contigs: 364

Minimum length: 100

Maximum length: 50093

Total length: 2069108

Mean length: 5684

N50 contig number: 44

N50 contig length (L50): 15404

N90 contig number: 146

N90 contig length (L50): 3907

N95 contig number: 180

N95 contig length (L50): 2305

Action

Upload

## 6.16.2 Select from isolate query

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

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



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

The screenshot shows the 'Isolate query/update' form. At the top, there are navigation links: 'PubMLST', 'Database home', 'Curator home', and 'Contents'. Below this, a status bar indicates 'Logged in: Keith Jolley (keith)' with links for 'Log out' and 'Change password'. The main heading is 'Isolate query/update'. Below the heading, there's a search bar with the text 'Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.' The search criteria are set to 'Isolate provenance/phenotype fields' with 'isolate' in the dropdown, an equals sign, and 'fam18' in the text field. There are 'Display/sort options' with 'Order by: id' and 'ascending' selected. There are 'Action' buttons: 'Reset' and 'Submit'. Below the search bar, there's a section for '1 record returned. Click the hyperlink for detailed information.' with tabs for 'Delete', 'Tag scanning', and 'Projects'. The 'Delete' tab is active, showing 'Delete ALL', 'Scan', 'Select project...', and 'Link' buttons. Below this, there's a table with columns: 'Delete', 'Update', 'Sequence bin', 'New version', 'id', 'isolate', 'aliases', 'country', 'year', 'disease', 'species', 'serogroup', 'genogroup', 'capsule group', 'ST', 'MLST', 'Finotyping antigens'. The 'Sequence bin' column is highlighted with a red rectangular box. The table contains one record with the following data:

Delete	Update	Sequence bin	New version	id	isolate	aliases	country	year	disease	species	serogroup	genogroup	capsule group	ST	MLST	Finotyping antigens
				698	FAM18	NIBSC_3076; Z4259	USA	1983	invasive (unspecified/other)	Neisseria meningitidis	C		C	11	ST-11 complex	PorA VR1 5 ; PorA VR2 2 ; FetA VR F1-30

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 home Curator home Contents

Logged in: Keith Jolley (Keith) Log out Change password Toggle: 1

## Upload 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: ! Select sender ...

method:

run id:

assembly id:

Options

☐ Don't insert sequences shorter than 25 bps.

Link to experiment:

Alternatively upload FASTA file or enter Genbank accession

Select FASTA file:

Browse... No file selected.

Action

Reset Submit

### 6.16.3 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.

PubMLST Database home Curator home Contents

Logged in: Keith Jolley (Keith) Log out Change password Toggle: 1

## Upload 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: ! Select sender ...

method:

run id:

assembly id:

Options

☐ Don't insert sequences shorter than 25 bps.

Link to experiment:

Alternatively upload FASTA file or enter Genbank accession

Select FASTA file:

Browse... No file selected.

Action

Reset Submit

## 6.17 Batch uploading sequence contigs linked to multiple isolate records

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



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

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

 The screenshot shows the 'Batch upload sequence assemblies to multiple isolate records' form. At the top, there are navigation links: 'PubMLST', 'Database home', 'Curator home', and 'Contents'. Below this, a status bar indicates 'Logged in: Keith Jolley (Keith)' with links for 'Log out' and 'Change password'. The main heading is 'Batch upload sequence assemblies to multiple isolate records'. The form contains the following elements:
 


- A description: 'This function allows you to upload assembly contig files for multiple records together. The first step in the upload process is to state which assembly contig FASTA file should be linked to each isolate record. You can use any provenance metadata field that uniquely identifies an isolate. You can upload up to 100 genomes at a time.'
- An 'Identifying field name' section with a 'Field:' dropdown menu set to 'isolate'.
- A 'Filenames' section with instructions: 'Paste in tab-delimited text, e.g. copied from a spreadsheet, consisting of two columns. The first column should be the value for the isolate identifier field (specified above), and the second should be the filename that you are going to upload. You need to ensure that you use the full filename, including any suffix such as .fas or .fasta, which may be hidden by your operating system.'
- A text area containing the following tab-delimited text:
 

JB_21292	JB_21292.fasta
JB_21293	JB_21293.fasta
JB_21294	JB_21294.fasta
- An 'Action' section with 'Reset' and 'Submit' buttons.

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



PubMLST Database home Curator home Contents

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

### Batch upload sequence assemblies to multiple isolate records

Please upload the assembly contig files for each isolate record.

remove row	id	isolate	current sequence bin state		filename	upload status
			contigs	total size (bp)		
<input type="checkbox"/>	61222	JB_21292	-	-	JB_21292.fasta	✗
<input type="checkbox"/>	61223	JB_21293	-	-	JB_21293.fasta	✗
<input type="checkbox"/>	61224	JB_21294	-	-	JB_21294.fasta	✗

[Remove](#)

3 FASTA files left to upload.

Contig assembly files

Please upload contig assemblies with the filenames as you specified (indicated in the table). Individual filesize is limited to 64 MB. You can upload up to 64 MB in one go, although you can upload multiple times so that the total size of the upload can be larger.


Drop files here or click to upload.

[←](#) [Home](#)

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

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

PubMLST Database home Curator home Contents

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

### Batch upload sequence assemblies to multiple isolate records

Please upload the assembly contig files for each isolate record.

remove row	id	isolate	current sequence bin state		filename	upload status
			contigs	total size (bp)		
<input type="checkbox"/>	61222	JB_21292	-	-	JB_21292.fasta	✓
<input type="checkbox"/>	61223	JB_21293	-	-	JB_21293.fasta	✓
<input type="checkbox"/>	61224	JB_21294	-	-	JB_21294.fasta	✓

[Remove](#)

All files uploaded. The sequences have not yet been validated. This needs to be done before they can be added to the database.

Action

[Validate](#)

[←](#) [Home](#)

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

PubMLST Database home Curator home Contents

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

### Batch upload sequence assemblies to multiple isolate records

**Validation**

id	isolate	filename	valid FASTA	contigs	total size
61222	JB_21292	JB_21292.fasta	✓	364	2,069,108
61223	JB_21293	JB_21293.fasta	✓	1	2,194,961
61224	JB_21294	JB_21294.fasta	✓	1	2,272,360

You can upload 3 records.

**Attributes**

sender:

method:

**Options**

☐ Don't insert sequences shorter than 25 bps.

**Action**

You can also choose to filter out short contigs from the upload by selecting the checkbox and choosing the minimum length from the dropdown box in the options settings.

PubMLST Database home Curator home Contents

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

### Batch upload sequence assemblies to multiple isolate records

**Validation**

id	isolate	filename	valid FASTA	contigs	total size
61222	JB_21292	JB_21292.fasta	✓	364	2,069,108
61223	JB_21293	JB_21293.fasta	✓	1	2,194,961
61224	JB_21294	JB_21294.fasta	✓	1	2,272,360

You can upload 3 records.

**Attributes**

sender:

method:

**Options**

☒ Don't insert sequences shorter than 100 bps.

**Action**

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



## 6.18 Linking remote contigs to isolate records

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

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



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

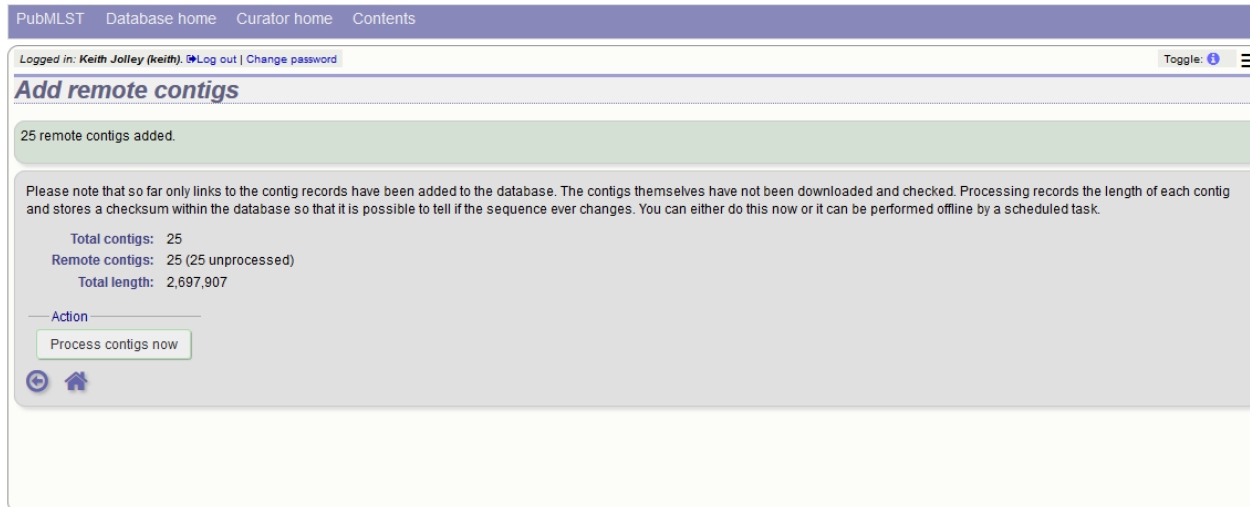
Press submit.

The screenshot shows the 'Add remote contigs' page. At the top, there is a navigation bar with links: PubMLST, Database home, Curator home, and Contents. Below this, a status bar indicates the user is logged in as 'Keith Jolley (keith)' with links for 'Log out' and 'Change password'. The page title is 'Add remote contigs'. A message states: 'This page allows you to link contigs stored in a remote BIGSdb database to an isolate record. Access to these contigs is via the BIGSdb RESTful API which must be running on the remote database. Valid URIs are in the form 'http://rest.pubmlst.org/db/[database\_config]/isolates/[isolate\_id]'.' Below this, there is a form section titled 'Enter details'. It contains two input fields: 'isolate id: !' with the value '1' and 'isolate record URI: !' with the value 'http://rest.pubmlst.org/db/pubmlst\_rmlst\_isolates/isolates/933'. To the right of these fields is an 'Action' section with 'Reset' and 'Submit' buttons. The 'Submit' button is highlighted with a red box.

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

The screenshot shows the 'Add remote contigs' page after the 'Submit' button was clicked. The page title remains 'Add remote contigs'. A message states: 'Checking contigs'. Below this, it says 'Downloading isolate record ... done.' and displays the following information: 'Contigs: 25' and 'Total length: 2,697,907 bp'. Below this information is an 'Action' section with an 'Upload' button, which is highlighted with a red box.

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



PubMLST Database home Curator home Contents

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

### Add remote contigs

25 remote contigs added.

Please note that so far only links to the contig records have been added to the database. The contigs themselves have not been downloaded and checked. Processing records the length of each contig and stores a checksum within the database so that it is possible to tell if the sequence ever changes. You can either do this now or it can be performed offline by a scheduled task.

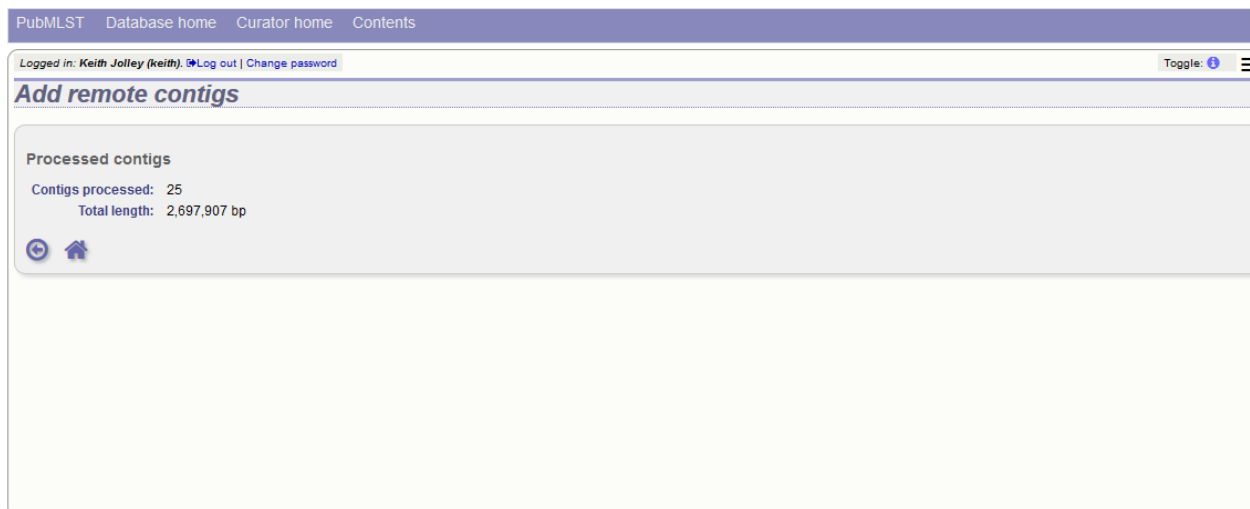
Total contigs: 25  
 Remote contigs: 25 (25 unprocessed)  
 Total length: 2,697,907

Action

Process contigs now

⏮ ⏭

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



PubMLST Database home Curator home Contents

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

### Add remote contigs

Processed contigs

Contigs processed: 25  
 Total length: 2,697,907 bp

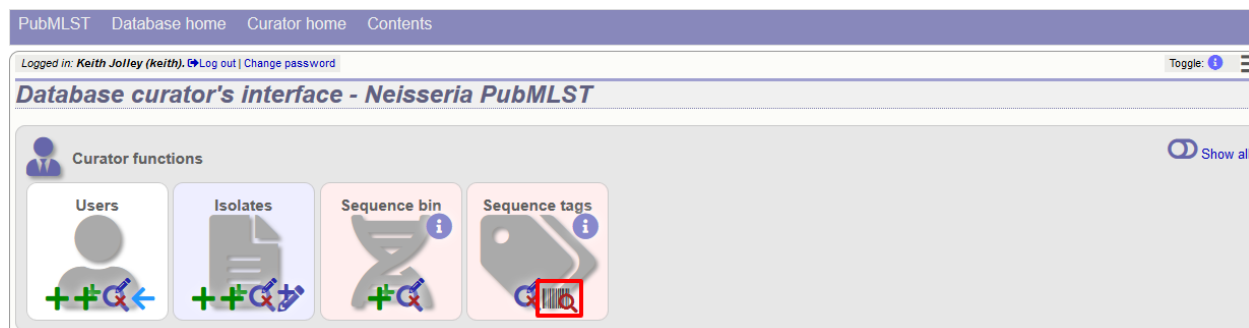
⏮ ⏭

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

## 6.19 Automated web-based sequence tagging

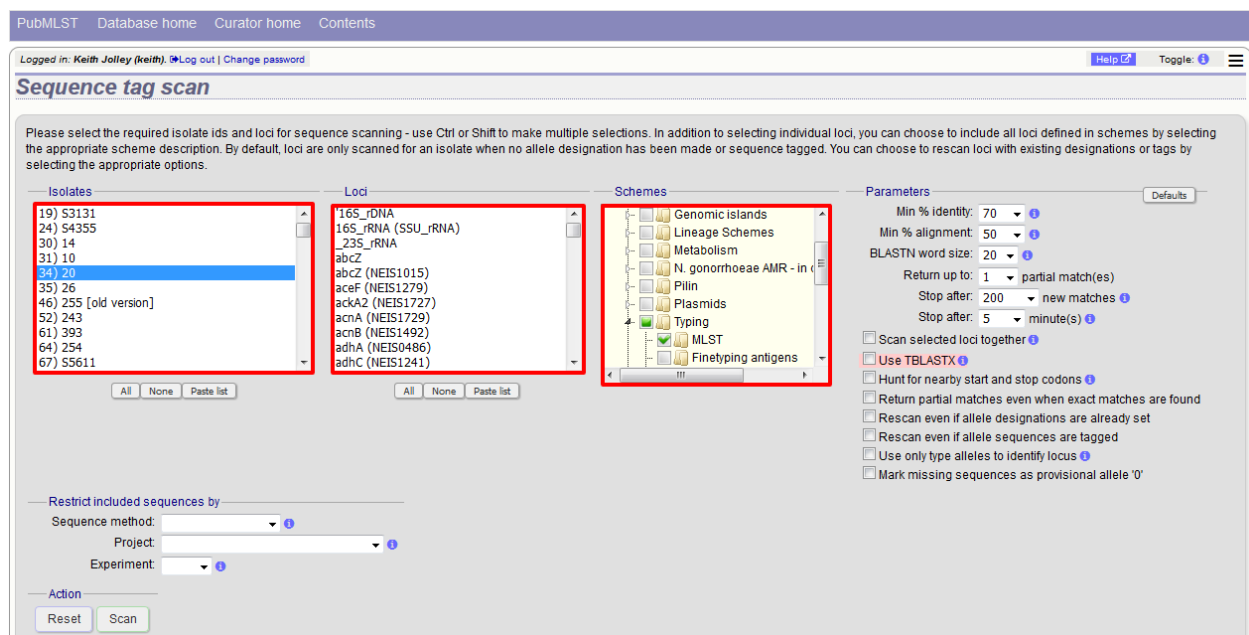
Sequence tagging, or tag-scanning, is the process of identifying alleles by scanning the sequence bin linked to an isolate record. Defined loci can either have a single reference sequence, that is defined in the locus table, or they can be linked to an external database that contains the sequences for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

Select ‘scan’ sequence tags on the curator’s index page.



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

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



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

By default, for each isolate only loci that have not had either an allele designation made or a sequence region scanned will be scanned again. To rescan in these cases, select either or both the following:

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

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

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

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

PubMLST Database home Curator home Contents

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

### 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
- 34) 20
- 35) 26
- 46) 255 [old version]
- 52) 243
- 61) 393
- 64) 254
- 67) S5611

All None Paste list

Loci

- 16S\_rDNA
- 16S\_rRNA (SSU\_rRNA)
- 23S\_rRNA
- abcZ
- abcZ (NEIS1015)
- aceF (NEIS1279)
- ackA2 (NEIS1727)
- acnA (NEIS1729)
- acnB (NEIS1492)
- adhA (NEIS0486)
- adhC (NEIS1241)

All None Paste list

Schemes

- Genomic islands
- Lineage Schemes
- Metabolism
- N. gonorrhoeae AMR - in
- Pilin
- Plasmids
- 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)

☐ Scan selected loci together

☒ 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

☐ Use only type alleles to identify locus

☐ 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). Alternatively, if 'Scan selected loci together' is available and selected, it may take longer to return initial results but total time should be less (e.g. a 2000 loci cgMLST scheme may be returned in 1-2 minutes). Any identified sequences will be listed in a table, with checkboxes indicating whether allele sequences or sequence regions are to be tagged.

PubMLST 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

### 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	fumC	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

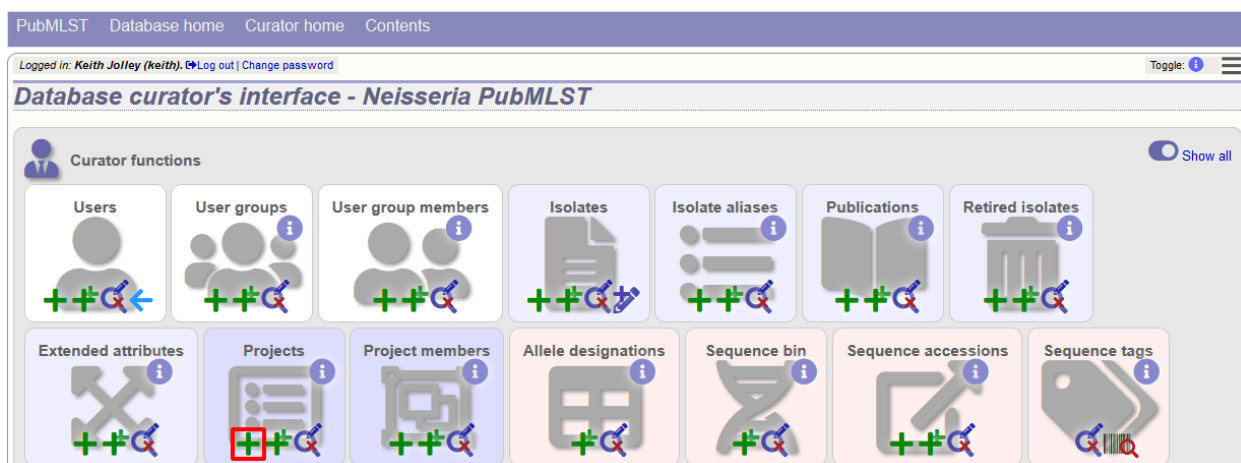
*Automated offline sequence tagging*

## 6.20 Projects

### 6.20.1 Creating the project

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

Click the add (+) project link on the curator's main page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.



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 be displayed at the top of an isolate information page (but see 'isolate\_display' flag below). The full description can include HTML formatting, including image links.

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

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

There are a further two option flags:

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

Click 'Submit'.



PubMLST Database home Curator home Contents

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

## Add new project description

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

Record

id: 3

short description: MRF Meningococcus Genome Library

isolate display: ☒ true ☐ false

list: ☒ true ☐ false

private: ☐ true ☒ false

no quota: ☒ true ☐ false

curator: Keith Jolley (keith)

timestamp: 2018-06-08

full description: 

```
<div style="float:right; padding: 0 2em"></div>
<div><p>The <a href="http://www.meningitis.org/research/genome">MRF Meningococcus Genome Library</a> is a collaboration between Public Health England, the Scottish Haemophilus, Legionella, Meningococcus and Pneumococcus Reference (SHLMPB) Laboratory, The Wellcome Trust Sanger Institute and the University of Oxford, funded by the Meningitis Research
```

Action

Reset Submit

## 6.20.2 Explicitly adding isolates to a project

Explicitly adding isolates to the project can be done individually or in batch mode. To add individually, click the add (+) project member link on the curator's main page. This function is normally hidden, so you may need to click the 'Show all' toggle to display it.

PubMLST Database home Curator home Contents

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

## Database curator's interface - Neisseria PubMLST

Curator functions Show all


Users User groups User group members Isolates Isolate aliases Publications Retired isolates

Extended attributes Projects Project members Allele designations Sequence bin Sequence accessions

Sequence tags Experiments Experiment contigs

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


PubMLST Database home Curator home Contents

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	Action
project id: 3) MRF Meningococcus Genome Library	Reset Submit
isolate id: 18968	
curator: Keith Jolley (keith)	
timestamp: 2018-06-08	


















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

PubMLST Database home Curator home Contents

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

### Database curator's interface - Neisseria PubMLST

 Curator functions 

<b>Users</b>  ++G	<b>User groups</b>  ++G	<b>User group members</b>  ++G	<b>Isolates</b>  ++G	<b>Isolate aliases</b>  ++G	<b>Publications</b>  ++G	<b>Retired isolates</b>  ++G
<b>Extended attributes</b>  ++G	<b>Projects</b>  ++G	<b>Project members</b>  ++G	<b>Allele designations</b>  +G	<b>Sequence bin</b>  +G	<b>Sequence accessions</b>  ++G	

Download an Excel submission template:

PubMLST Database home Curator home Contents

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


**Batch insert project members**

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

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

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

Action  
Reset Submit



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 home Curator home Contents

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

**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	19033
3	19034

Action  
Reset **Submit**



## 6.21 Isolate record versioning

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

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

Links to different versions are displayed within an isolate record:

PubMLST Database home Contents

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

### Full information on isolate M17661 (id:19377)

**Provenance/meta data**

id: 19377	species: <i>Neisseria meningitidis</i>
isolate: M17661	serogroup: W
strain designation: W: P1.5-1,10-8: F-ND: ST-11 (cc11)	capsule group: W
country: USA	sender: Keith Jolley, University of Oxford, UK
continent: North America	curator: Auto Tagger
region: MI	update history: <a href="#">51 updates</a> <a href="#">show details</a>
year: 2008	date entered: 2012-06-27
disease: invasive (unspecified/other)	datestamp: 2018-01-31

**Versions**

More than one version of this isolate record exist.

Newer versions: [40500](#)

**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: 2,511	N50 contig number: 591	N95 contig number: 2,006
total length: 1,805,445 bp	N50 length (L50): 1,011	N95 length (L95): 267
max length: 4,833 bp	N90 contig number: 1,720	loci tagged: 1,305
mean length: 720 bp	N90 length (L90): 366	detailed breakdown: <a href="#">Display</a>

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:

PubMLST Database home Curator home Contents

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

Toggle:

## Database curator's interface - Neisseria PubMLST

**Curator functions**

Users

Isolates

Sequence bin

Sequence tags

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

PubMLST Database home Curator home Contents

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

Toggle:

## Isolate query/update

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields:  =  Enter value...

Display/sort options: Order by:  ascending Display: 25 records per page

Action:

58,801 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Select project...

Page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [Last](#)

Delete	Update	Sequence bin	New version	id	isolate	aliases	country	year	Isolate fields	disease	species	serogroup/genogroup	capsule group	ST	MLST	clonal complex	PortA VR1	PortA VR2	FeiA VR
				1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)		Neisseria meningitidis	A	A	4	ST-4 complex	5-2	10	F1-5	
				2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicemia		Neisseria meningitidis	A	A	1	ST-1 complex	5-2	10	F5-1	
				3	M00242905		UK	2000	invasive (unspecified/other)		Neisseria meningitidis	B	B	1099		19	15		
				4	M1027	B43; NIBSC_3076; Z1043	USA	1937	invasive (unspecified/other)		Neisseria meningitidis	A	A	4	ST-4 complex				
				5	M00240227		UK	2000	invasive (unspecified/other)		Neisseria meningitidis	B	B	1100	ST-32 complex	7	16		
				6	M00282207		UK	2000	invasive (unspecified/other)		Neisseria meningitidis	W	W	1101	ST-22 complex				
				7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)		Neisseria meningitidis	A	A	5	ST-5 complex	20	9	F3-1	
				8	M00242007		UK	2000	invasive (unspecified/other)		Neisseria meningitidis	B	B	1102	ST-18 complex		14		
				9	0021/84		Czech Republic	1984	invasive (unspecified/other)		Neisseria meningitidis	W	W	114	ST-22 complex				
				10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)		Neisseria meningitidis	A	A	1	ST-1 complex	18-1	3	F5-1	
				11	129C	B92; NIBSC_2828; Z1092	Germany	1964	invasive (unspecified/other)		Neisseria meningitidis	A	A	1	ST-1 complex	5-2	10	F3-6	
				12	0090/89		Czech Republic	1989	invasive (unspecified/other)		Neisseria meningitidis	B	B	1015	ST-32 complex	7	16		
				13	139M	B99; NIBSC_2795; Z1099	Philippines	1968			Neisseria meningitidis	A	A	1	ST-1 complex	5-2	10	F5-1	
				14	0120/95		Czech Republic	1995	invasive (unspecified/other)		Neisseria meningitidis	X	X	117			14		
				15	1		Germany	1999	carrier		Neisseria meningitidis	E	E	864					

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.


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

## 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



☒ Add new version to projects


 **Provenance/meta data**

id: 7  
isolate: 7891  
aliases: B54; NIBSC\_2760; Z1054  
strain designation: A: P1.20.9; F3-1: ST-5 (cc5)  
country: Finland  
continent: Europe  
year: 1975  
disease: invasive (unspecified/other)  
source: CSF  
epidemiology: epidemic  
species: *Neisseria meningitidis*

serogroup: A  
capsule group: A  
MLEE designation: Subgroup III  
serotype: 4,21  
sero subtype: P1.9  
ET no: 48  
ENA accession: ERS006946 [www.ebi.ac.uk](http://www.ebi.ac.uk)  
comments: Pili I,IIa  
sender: Wendell Zollinger, Dept Bacterial Diseases, Walter Reed Army Institute of Research, Washington DC, USA

curator: Nina Billows, University of Oxford (E-mail: [nina.billows@some.ox.ac.uk](mailto:nina.billows@some.ox.ac.uk))  
update history: [102 updates](#) | [show details](#)  
date entered: 2001-02-07  
timestamp: 2018-06-04

 **Publications (8)** 

 **Sequence bin**

contigs: 199  
total length: 2,057,385 bp  
max length: 112,831 bp

mean length: 10,339 bp  
N50 contig number: 29  
N50 length (L50): 23,361


N90 contig number: 94  
N90 length (L90): 6,218  
N95 contig number: 116

N95 length (L95): 3,616  
loci tagged: 2,185  
detailed breakdown: [Display](#)

## CHAPTER 7

### 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 [Help](#)

## Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- alleles
- MLST profiles

New allele sequence submissions waiting for curation

Your account is authorized to handle the following submissions:

Submission id	Submitted	Updated	Submitter	Locus	Sequences
BIGSdb_20150709132553_20864_89729	2015-07-09	2015-07-09	Joe Bloggs	NEIS0001 (lpxC)	3

[Return to index page](#)

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



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

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

### Curate submission

Submission: BIGSdb\_20150709132553\_20864\_89729

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-07-09

status: pending

locus: NEIS0001 (lpxC)

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	ATGCTGCAAGAAGACTTCGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>
UK323	924	ATGCTGCAAGAAGACTTCGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>
UK347	924	ATGCTGCAAGAAGACTTCGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>

Batch curate

Update

Messages

Archive

Archive of submission and any supporting files:

Download

Message:

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 | Help

### Curate submission

Submission: BIGSdb\_20150709132553\_20864\_89729

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-07-09

status: pending

locus: NEIS0001 (lpxC)

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	ATGCTGCAAGAAGACTTCGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>
UK323	924	ATGCTGCAAGAAGACTTCGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>
UK347	924	ATGCTGCAAGAAGACTTCGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>

Batch curate

Update

Messages

Archive

Archive of submission and any supporting files:

Download

Message:

## 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\_20150709132553\_20864\_89729

Summary

type: alleles  
 submitter: Joe Bloggs, University of Oxford, UK  
 datestamp: 2015-07-09  
 status: pending  
 locus: NEIS0001 (lpxC)  
 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	ATGCTGCAAGAAGAACTTCGGC ... ACATCAGAAATTTGGGATAA	✓	pending	<a href="#">Curate</a>
UK323	924	ATGCTGCAAGAAGAACTTTGGC ... ACATCAGAAATTTGGGATAA	✓	pending	<a href="#">Curate</a>
UK347	924	ATGCTGCAAGAAGAACTTTGGC ... ACATCAGAAATTTGGGATAA	✓	pending	<a href="#">Curate</a>

Batch curate

Messages

Archive

Archive of submission and any supporting files:

Download

Message: Append Send now

Update

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

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 allele sequence

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

Record

locus!: NEIS0001

allele id!: 210

sequence!: ATGCTGCAAGAAGAACTTCGGCGAAATCCATCAGCGTTACCGGAGTCGGCCTGCATTCCGGCGAACGGGTGCGACTG  
 ACCCTGCACCCCGCGCCTGAAAACAGCGGGATTTCCTTCGGCCGTACCGATTGGACGGCGAGATGGGCGAACAA  
 ATCAAGCTGACCCCTTATTGATCAACGATACCGCGCTTTCCTCCACCATCGTTACCGACAAAGCGGTGCGCGTC  
 GGCACGATCGAACACATTATGTCGGCGCTGTCCGCCCTACGGTATCGACAACGCGCTGATTGAGCTGAACGCACCC  
 GAAATCCCGATTATGGACGGCTCCAGCCTGCCGTTTATTACCTTTTGCAAGATGCGGGCGTGGTGCATCAAAAG  
 GCGCAAAAGCGTTTTTTGAAAATCCTCAAGCCTGTGCAAAATCAAAGAAGCGGGCAAATGGGTGCGCTTTACGCCG  
 TATGACGGCTTTAAGGTAACGCTGACCATCGAATTGACCATCCGGCTTTCAACCGCAGCTCGCCCACTTTTGAA

status!: unchecked

sender!: Bloggs, Joe (jbloggs)

curator!: Keith Jolley (keith)

date entered!: 2015-07-09

datestamp!: 2015-07-09

comments:

Flags: atypical  
 contains IS element  
 downstream fusion

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

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

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

## Add new allele sequence

Sequence NEIS0001 (lpxC): 210 added!

[Add another](#) [Return to submission](#) [Back to main page](#)

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

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

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

## Curate submission

Submission: BIGSdb\_20150709132553\_20864\_89729

**Summary**

type: alleles  
 submitter: Joe Bloggs, University of Oxford, UK  
 datestamp: 2015-07-09  
 status: pending  
 locus: NEIS0001 (lpxC)  
 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	ATGCTGCAAGAACTTGGC ... ACATCAGAAATTGTGGATAA	✓	assigned	210
UK323	924	ATGCTGCAAGAACTTGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>
UK347	924	ATGCTGCAAGAACTTGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>

[Batch curate](#) [Update](#)

**Messages**

Message:  [Append](#) [Send now](#)

**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.

**Note:** Batch curation is only available for loci that do not have extended attributes defined. Entries for these loci

require additional values set for these additional fields and so need to be handled individually.

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\_20150709132553\_20864\_89729

**Summary**

- type: alleles
- submitter: Joe Bloggs, University of Oxford, UK
- datestamp: 2015-07-09
- status: pending
- locus: NEIS0001 (lpxC)
- 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	ATGCTGCAAAGAAGCTTCGGC ... ACATCAGAAATTGTGGATAA	✓	assigned	210
UK323	924	ATGCTGCAAAGAAGCTTCGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>
UK347	924	ATGCTGCAAAGAAGCTTCGGC ... ACATCAGAAATTGTGGATAA	✓	pending	<a href="#">Curate</a>

**Batch curate** (highlighted)

**Messages**

Message:

**Archive**

Archive of submission and any supporting files:  
[Download TAR](#)

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

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

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 (lpxC)
- status: unchecked
- sender: Bloggs, Joe (jbloggs)

**sequence (FASTA):**

```
>UK323
ATGCTGCAAAGAAGCTTCGGCAGAAATCGATCAGCGTTACCGGAGTCGGGCTGCATTCCGGGC
GAACGCGTCGCGCTGACCCCTGCACCCGCGCCCTGAAAACAGCGGGATTTCCTTCGCCCTG
ACCGATTGGACGGCGAGATGGCGAACAATCAAGCTGACCCCTTATTGATCAACGAT
ACCGCCCTTCTCCACCATCGTTACCGAACAAGCGGTGCGCGTCGGCAGATCGAACAC
ATTATGTCGCGCTGTCGCGCTACGGCATCGACAACGCGCTGATTGAGCTGAACGCGCC
GAAATCCCGATTATGACGCGCTCCAGCCTGCCGTTTATTACCTTTTCAAGATCGGGC
GTGGTCGATCAAAAGGCGCAAAAGCGTTTTTTGAAAATCCTCAAGCCTGCGAAATCAA
GAAGCGGCAAAATGGGTGCGCTTTACGCGGTATGACGCTTTAAGGTAAACGCTGACCATC
GAATTCGACCATCGGTTTTCAACCGAGCTCGCCACTTTTGAATCGATTTCGCGCGC
AAATCCTACATCGACGAAATCGCGCGCGCGCACTTTCGGCTTTATGCACGAAGTGGAA
```

☒ 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**

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	211	OK	<a href="#">Upload valid sequences</a>
UK347	212	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.

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\_20150709132553\_20864\_89729

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-07-09

status: pending

locus: NEIS0001 (lpxC)

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	ATGCTGCAAGAAGACTTCGGC ... ACATCAGAAATTGGGATAA	✓	assigned	210
UK323	924	ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGGGATAA	✓	assigned	211
UK347	924	ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGGGATAA	✓	assigned	212

Messages

Archive

Archive of submission and any supporting files:

Download

Message:

Action

### 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.

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\_20150709132553\_20864\_89729

Summary

type: alleles

submitter: Joe Bloggs, University of Oxford, UK

datestamp: 2015-07-09

status: pending

outcome: accepted - data uploaded

locus: NEIS0001 (lpxC)

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	ATGCTGCAAGAAGACTTCGGC ... ACATCAGAAATTGGGATAA	✓	rejected	
UK323	924	ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGGGATAA	✓	rejected	
UK347	924	ATGCTGCAAGAAGACTTTGGC ... ACATCAGAAATTGGGATAA	✓	rejected	

Batch curate

Messages

Archive

Archive of submission and any supporting files:

Download

Message:

### 7.1.4 Requesting additional information

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

### 7.1.5 Closing the submission

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


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

Logged in: Keith Jolley (keith) Log out | Change password Help C?

### Curate submission

Submission: BIGSdb\_20150709132553\_20864\_89729

Summary

type: alleles  
 submitter: Joe Bloggs, University of Oxford, UK  
 datestamp: 2015-07-09  
 status: pending  
 locus: NEIS0001 (lpxC)  
 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	ATGCTGCAAAGAAGCTTCGGC ... ACATCAGAAATTGTGGATAA	✓	rejected	
UK323	924	ATGCTGCAAAGAAGCTTCGGC ... ACATCAGAAATTGTGGATAA	✓	rejected	
UK347	924	ATGCTGCAAAGAAGCTTCGGC ... ACATCAGAAATTGTGGATAA	✓	rejected	

Update

Messages

Timestamp	User	Message
2015-07-09 12:39:35+00	Keith Jolley	These sequences are not NEIS0001 alleles

Archive

Archive of submission and any supporting files:  
 Download 

Action

Message: Append Send now **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 Help C?

### Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- alleles
- MLST profiles

New allelic profile submissions waiting for curation

Your account is authorized to handle the following submissions:

Submission id	Submitted	Updated	Submitter	Scheme	Profiles
<b>BIGSdb_20150709134405_4219_35535</b>	2015-07-09	2015-07-09	Joe Bloggs	MLST	3

Return to index page

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

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

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

### Curate submission


Submission: BIGSdb\_20150709134405\_4219\_35535

Summary

type: profiles  
 submitter: Joe Bloggs, University of Oxford, UK  
 datestamp: 2015-07-09  
 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	<a href="#">Curate</a>
UK33	7	56	4	3	2	12	12	pending	<a href="#">Curate</a>
UK34	76	3	5	3	87	43	34	pending	<a href="#">Curate</a>

Batch curate

Messages

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.

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\_20150709134405\_4219\_35535

Summary

type: profiles  
 submitter: Joe Bloggs, University of Oxford, UK  
 datestamp: 2015-07-09  
 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	<a href="#">Curate</a>
UK33	7	56	4	3	2	12	12	pending	<a href="#">Curate</a>
UK34	76	3	5	3	87	43	34	pending	<a href="#">Curate</a>


Batch curate

Messages

Message:

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: [i](#)

## Add new MLST profile

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

**Record**

ST: !

adk: !

abcZ: !

aroE: !

fumC: !

gdh: !

pdhC: !

pgm: !

sender: !

clonal\_complex:

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

date\_entered: ! **2015-07-09**

datestamp: ! **2015-07-09**

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-10056 added!

[Return to submission](#) [Add another](#) | [Back to main page](#)

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

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

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

## Curate submission

Submission: BIGSdb\_20150709134405\_4219\_35535

**Summary**

type: profiles  
 submitter: Joe Bloggs, University of Oxford, UK  
 datestamp: 2015-07-09  
 status: pending

**Profiles**


Identifier	adk	abcZ	aroE	fumC	gdh	pdhC	pgm	Status	Assigned ST
UK32	43	2	12	32	32	3	2	assigned	10056
UK33	7	56	4	3	2	12	12	pending	<a href="#">Curate</a>
UK34	76	3	5	3	87	43	34	pending	<a href="#">Curate</a>

[Batch curate](#) [Update](#)

**Messages**

Archive

Archive of submission and any supporting files:

[Download](#) 

Message:  [Append](#) [Send now](#)

## 7.2.2 Batch profile curation

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

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

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

### Curate submission

Submission: BIGSdb\_20150709134405\_4219\_35535

**Summary**

type: profiles  
submitter: [Joe Bloggs](#), University of Oxford, UK  
datestamp: 2015-07-09  
status: pending

**Profiles**

Identifier	adk	abcZ	aroE	fumC	gdh	pdhC	pgm	Status	Assigned ST
UK32	43	2	12	32	32	3	2	assigned	10056
UK33	7	56	4	3	2	12	12	pending	<a href="#">Curate</a>
UK34	76	3	5	3	87	43	34	pending	<a href="#">Curate</a>

[Batch curate](#) [Update](#)

**Messages**

**Archive**  
Archive of submission and any supporting files:  
[Download](#)

Message: [Append](#) [Send now](#)

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

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

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](#) [Help](#)

## Curate submission

Submission: BIGSdb\_20150709134405\_4219\_35535

**Summary**

type: profiles  
submitter: Joe Bloggs, University of Oxford, UK  
datestamp: 2015-07-09  
status: pending

**Profiles**

Identifier	adk	abcZ	aroE	fumC	gdh	pdhC	pgm	Status	Assigned ST
UK32	43	2	12	32	32	3	2	assigned	10056
UK33	7	56	4	3	2	12	12	assigned	10057
UK34	76	3	5	3	87	43	34	assigned	10058

**Messages**

Message:

**Archive**

Archive of submission and any supporting files:

[Download](#)

**Action**

### 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](#) [Help](#)

## Curate submission

Submission: BIGSdb\_20150709134405\_4219\_35535

**Summary**

type: profiles  
submitter: Joe Bloggs, University of Oxford, UK  
datestamp: 2015-07-09  
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	<a href="#">Curate</a>
UK33	7	56	4	3	2	12	12	rejected	<a href="#">Curate</a>
UK34	76	3	5	3	87	43	34	rejected	<a href="#">Curate</a>

**Messages**

Message:

**Archive**

Archive of submission and any supporting files:

[Download](#)

## 7.2.4 Requesting additional information

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

## 7.2.5 Closing the submission

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

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

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

### Curate submission

Submission: BIGSdb\_20150709134405\_4219\_35535

**Summary**

type: profiles  
submitter: Joe Bloggs, University of Oxford, UK  
timestamp: 2015-07-09  
status: pending

**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
2015-07-09 12:55:23+00	Keith Jolley	You need to submit some representative isolates data for these profiles.

Message:  Append Send now **Close submission**

**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 Help

### Manage submissions

**Submit new data**  
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

**Submission type:**

- isolates

**Pending submissions**  
You have submitted the following submissions that are pending curation:

Submission id	Submitted	Updated	Type	Details
BIGSdb_20150709121747_1342_99624	2015-07-09	2015-07-09	isolates	2 isolates

[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 Help

### Curate submission

Submission: BIGSdb\_20150709121747\_1342\_99624

Summary

type: isolates  
submitter: Joe Bloggs, University of Oxford, UK  
datestamp: 2015-07-09  
status: pending

Isolates

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

Batch curate Record status: pending Update

Messages Archive  
Archive of submission and any supporting files:  
Download TAR

Message: Append Send now

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 Help

### Curate submission

Submission: BIGSdb\_20150709121747\_1342\_99624

Summary

type: isolates  
submitter: Joe Bloggs, University of Oxford, UK  
datestamp: 2015-07-09  
status: pending

Isolates

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2


Batch curate Record status: pending Update


Messages Archive  
Archive of submission and any supporting files:  
Download TAR

Message: Append Send now

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


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


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

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

### Batch insert isolates

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

- 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:

Bloggs, Joe (jbloggs)
Value will be overridden if you include a sender field in your pasted data.

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

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
F1-5	5	2						6
UK322	UK	2014	meningitis		blood	Neisseria meningitidis	B	
2	3	17	3	8	4	6	F1-5	5
							2-1	

Reset
Submit

[Back](#)

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

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](#)

**Note:** Depending on the database policy, definitions of new scheme profiles, e.g. for MLST, may require submission



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

### 7.3.1 Requesting additional information

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

### 7.3.2 Closing the submission

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

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

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

## Curate submission

Submission: BIGSdb\_20150709121747\_1342\_99624

**Summary**

type: isolates  
 submitter: Joe Bloggs, University of Oxford, UK  
 datestamp: 2015-07-09  
 status: pending

**Isolates**

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

[Batch curate](#)

**Record status:** accepted [Update](#)


**Messages**

Archive of submission and any supporting files:

[Download](#)

Message: [Append](#) [Send now](#)

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


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

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

## Curate submission

Submission: BIGSdb\_20150709121747\_1342\_99624

**Summary**

type: isolates  
 submitter: [Joe Bloggs](#), University of Oxford, UK  
 datestamp: 2015-07-09  
 status: pending


**Isolates**

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

Record status: accepted Update

**Messages**

Archive of submission and any supporting files: Close submission

[Download](#) 

Message: Append Send now

## 8.1 Automated offline sequence tagging

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

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

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

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

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

```
autotag.pl --database <database configuration>
```

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

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

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

A full list of options can be found by typing:

```
autotag.pl --help

NAME
    autotag.pl - BIGSdb automated allele tagger
```

(continues on next page)

(continued from previous page)

## SYNOPSIS

```
autotag.pl --database NAME [options]
```

## OPTIONS

```
-0, --missing
    Marks missing loci as provisional allele 0. Sets default word size to 15.

-d, --database NAME
    Database configuration name.

-e, --exemplar
    Only use alleles with the 'exemplar' flag set in BLAST searches to identify
    locus within genome. Specific allele is then identified using a database
    lookup. This may be quicker than using all alleles for the BLAST search,
    but will be at the expense of sensitivity. If no exemplar alleles are set
for a locus then all alleles will be used. Sets default word size to 15.

-f --fast
    Perform single BLAST query against all selected loci together. This will
    take longer to return any results but the overall scan should finish
    quicker. This method will also use more memory - this can be used with
    --exemplar to mitigate against this.

-h, --help
    This help page.

-i, --isolates LIST
    Comma-separated list of isolate ids to scan (ignored if -p used).

--isolate_list_file FILE
    File containing list of isolate ids (ignored if -i or -p used).

-I, --exclude_isolates LIST
    Comma-separated list of isolate ids to ignore.

-l, --loci LIST
    Comma-separated list of loci to scan (ignored if -s used).

-L, --exclude_loci LIST
    Comma-separated list of loci to exclude

-m, --min_size SIZE
    Minimum size of seqbin (bp) - limit search to isolates with at least this
    much sequence.

-n, --new_only
    New (previously untagged) isolates only. Combine with --new_max_alleles
if required.

--new_max_alleles ALLELES
    Set the maximum number of alleles that can be designated or sequences
    tagged before an isolate is not considered new when using the --new_only
    option.

-o, --order
    Order so that isolates last tagged the longest time ago get scanned first
```

(continues on next page)

(continued from previous page)

```

    (ignored if -r used).

--only_already_tagged
    Only check loci that already have a tag present (but no allele designation).
    This must be combined with the --already_tagged option or no loci will
    match. This option is used to perform a catch-up scan where a curator has
    previously tagged sequence regions prior to alleles being defined, without
    the need to scan all missing loci.

-p, --projects LIST
    Comma-separated list of project isolates to scan.

-P, --exclude_projects LIST
    Comma-separated list of projects whose isolates will be excluded.

-q, --quiet
    Only error messages displayed.

-r, --random
    Shuffle order of isolate ids to scan.

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

-R, --locus_regex REGEX
    Regex for locus names.

-s, --schemes LIST
    Comma-separated list of scheme loci to scan.

-t, --time MINS
    Stop after t minutes.

--threads THREADS
    Maximum number of threads to use.

-T, --already_tagged
    Scan even when sequence tagged (no designation).

-v, --view VIEW
    Isolate database view (overrides value set in config.xml).

-w, --word_size SIZE
    BLASTN word size.

-x, --min ID
    Minimum isolate id.

-y, --max ID
    Maximum isolate id.

```

## 8.2 Defining exemplar alleles

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

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

There is a script called 'find\_exemplars.pl' in the BIGSdb `scripts/maintenance` directory.

A full list of options can be found by typing:

```
find_exemplars.pl --help

NAME
    find_exemplars.pl - Identify and mark exemplar alleles for use
    by tagging functions

SYNOPSIS
    find_exemplars.pl --database NAME    [options]

OPTIONS

--database NAME
    Database configuration name.

--datatype DNA|peptide
    Only define exemplars for specified data type (DNA or peptide)

--exclude_loci LIST
    Comma-separated list of loci to exclude

--help
    This help page.

--loci LIST
    Comma-separated list of loci to scan (ignored if -s used).

--locus_regex REGEX
    Regex for locus names.

--schemes LIST
    Comma-separated list of scheme loci to scan.

--update
    Update exemplar flags in database.

--variation DISSIMILARITY
    Value for percentage identity variation that exemplar alleles
    cover (smaller value will result in more exemplars). Default: 10.
```

## 8.3 Automated offline allele definition

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

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

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

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

```
scannew.pl --database <database configuration>
```

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

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

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

A full list of options can be found by typing:

```
scannew.pl --help

NAME
    scannew.pl - BIGSdb automated allele definer

SYNOPSIS
    scannew.pl --database NAME [options]

OPTIONS
-a, --assign
    Assign new alleles in definitions database.

--allow_frameshift
    Allow sequences to contain a frameshift so that the length is not a
    multiple of 3, or an internal stop codon. To be used with
    --coding_sequences option to allow automated curation of pseudogenes.
    New alleles assigned will be flagged either 'frameshift' or 'internal stop
    codon' if appropriate. Essentially, combining these two options only
    checks that the sequence starts with a start codon and ends with a stop
    codon.

--allow_subsequences
    Allow definition of sub- or super-sequences. By default these will not
    be assigned.

-A, --alignment INT
    Percentage alignment (default: 100).

-B, --identity INT
    Percentage identity (default: 99).

-c, --coding_sequences
    Only return complete coding sequences.
```

(continues on next page)

(continued from previous page)

```
-d, --database NAME
    Database configuration name.

-h, --help
    This help page.

-i, --isolates LIST
    Comma-separated list of isolate ids to scan (ignored if -p used).

--isolate_list_file FILE
    File containing list of isolate ids (ignored if -i or -p used).

-I, --exclude_isolates LIST
    Comma-separated list of isolate ids to ignore.

-l, --loci LIST
    Comma-separated list of loci to scan (ignored if -s used).

-L, --exclude_loci LIST
    Comma-separated list of loci to exclude.

-m, --min_size SIZE
    Minimum size of seqbin (bp) - limit search to isolates with at least this
    much sequence.

-n, --new_only
    New (previously untagged) isolates only.

-o, --order
    Order so that isolates last tagged the longest time ago get scanned first
    (ignored if -r used).

-p, --projects LIST
    Comma-separated list of project isolates to scan.

-P, --exclude_projects LIST
    Comma-separated list of projects whose isolates will be excluded.

-r, --random
    Shuffle order of isolate ids to scan.

-R, --locus_regex REGEX
    Regex for locus names.

-s, --schemes LIST
    Comma-separated list of scheme loci to scan.

-t, --time MINS
    Stop after t minutes.

--threads THREADS
    Maximum number of threads to use.

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

(continues on next page)



(continued from previous page)

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

```
-T, --already_tagged
    Scan even when sequence tagged (no designation).

-v, --view VIEW
    Isolate database view (overrides value set in config.xml).

-w, --word_size SIZE
    BLASTN word size.

-x, --min ID
    Minimum isolate id.

-y, --max ID
    Maximum isolate id.
```

## 8.4 Cleanly interrupting offline curation

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

```
kill 23232
```

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

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

```
pkill autotag
```

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

Similarly, to terminate all allele autodefiner jobs, type

```
pkill scannew
```

## 8.5 Uploading contigs from the command line

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

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

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

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

A full list of options can be found by typing:

```
upload_contigs.pl --help

NAME
    upload_contigs.pl - Upload contigs to BIGSdb isolate database

SYNOPSIS
    upload_contigs.pl --database NAME --isolate ID --file FILE
                      --curator ID --sender ID [options]

OPTIONS
-a, --append
    Upload contigs even if isolate already has sequences in the bin.

-c, --curator ID
    Curator id number.

-d, --database NAME
    Database configuration name.

-f, --file FILE
    Full path and filename of contig file.

-h, --help
    This help page.

-i, --isolate ID
    Isolate id of record to upload to.

-m, --method METHOD
    Method, e.g. 'Illumina', default 'unknown'.

--min_length LENGTH
    Exclude contigs with length less than value.

-s, --sender ID
    Sender id number.
```

---

### Definition downloads

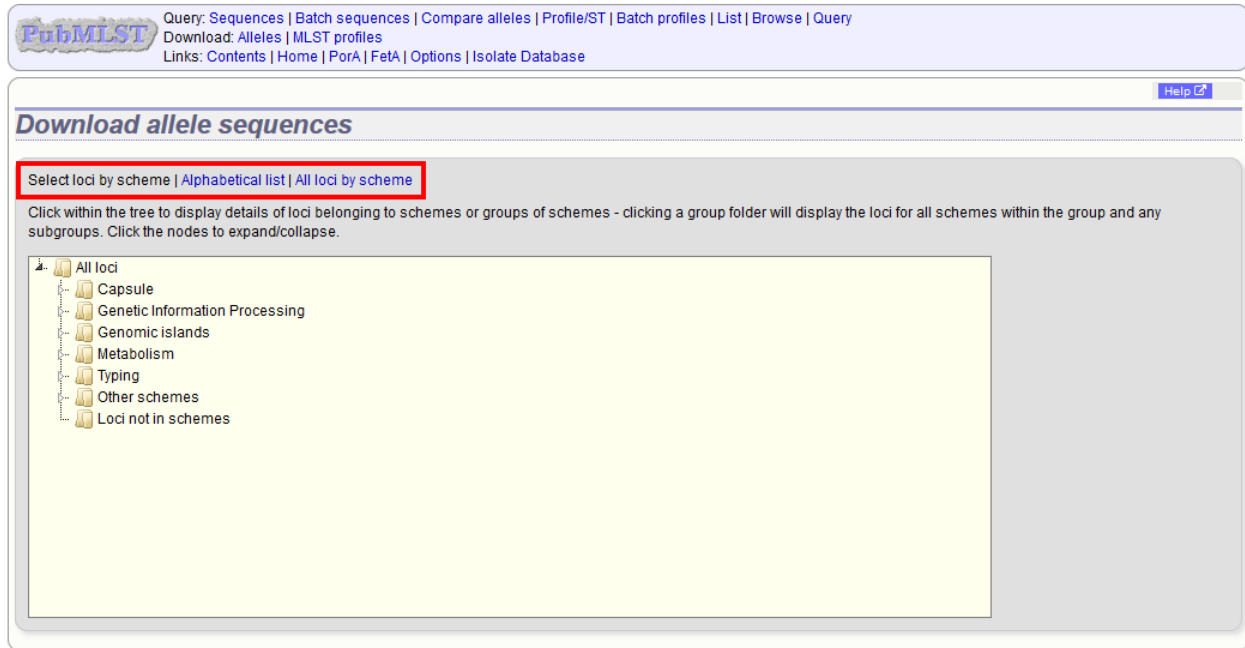
---

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

#### 9.1 Allele sequence definitions

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

### 9.1.1 Scheme tree



The screenshot shows the PubMLST website 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 "Download allele sequences". Below the heading, there are three tabs: "Select loci by scheme", "Alphabetical list", and "All loci by scheme". The "Select loci by scheme" tab is selected. Below the tabs, there is a text box that 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 the text box is a tree view showing the following structure:

- All loci
  - Capsule
  - Genetic Information Processing
  - Genomic Islands
  - Metabolism
  - Typing
  - Other schemes
  - 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**
      - Finetyping 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	<a href="#">Download</a>	DNA	728	Fixed: 433 bp		O. Harrison, K. Jolley	2015-06-22
adk	<a href="#">Download</a>	DNA	501	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-08
aroE	<a href="#">Download</a>	DNA	775	Fixed: 490 bp		O. Harrison, K. Jolley	2015-06-22
fumC	<a href="#">Download</a>	DNA	732	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-02
gdh	<a href="#">Download</a>	DNA	732	Fixed: 501 bp		O. Harrison, K. Jolley	2015-06-11
pdhC	<a href="#">Download</a>	DNA	747	Fixed: 480 bp		O. Harrison, K. Jolley	2015-06-25
pgm	<a href="#">Download</a>	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	<a href="#">Download</a>	DNA	728	Fixed: 433 bp		O. Harrison, K. Jolley	2015-06-22
adk	<a href="#">Download</a>	DNA	501	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-08
aroE	<a href="#">Download</a>	DNA	775	Fixed: 490 bp		O. Harrison, K. Jolley	2015-06-22
fumC	<a href="#">Download</a>	DNA	732	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-02
gdh	<a href="#">Download</a>	DNA	732	Fixed: 501 bp		O. Harrison, K. Jolley	2015-06-11
pdhC	<a href="#">Download</a>	DNA	747	Fixed: 480 bp		O. Harrison, K. Jolley	2015-06-25
pgm	<a href="#">Download</a>	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
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCCGCGCTGCGGT
TTGGGCGAAATCAAAATCCCCGAAAACGAGCCGGGTTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTTCGGCAACGACGTT
```

(continues on next page)

(continued from previous page)


```

ACCATCGGTATGGCGGGCGCGTCGGGCAATTTTCGAGCTGAACGCTCTATATGCCCGTCATC
GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCGTCGGCATTGAACCCGTACCGGAAAAAATCGACTATTTCTGCACCATTC
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAC
>fumC_2
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCCGCGCTGCGGT
TTGGGCGAAATCAAAATCCCCGAAAACGAGCCGGGTTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGCATGGCGGGCGCGTCGGGCAATTTTCGAGCTGAACGCTCTATATGCCCGTTATC
GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCATCGGCATCGAACCCTACCGGAAAAAATCGACTATTTCTGCACCATTC
CTGATGCTCGTTACCGCGTTAAACCGCAAAATCGGTTACGAAAAC
>fumC_3
GAAGCCTTGGGCGGACGCGATGCCGCCGTTGCCGCTTCGGGCGCATTGAAAACGCTGGCG
GCAAGCCTGAATAAAATCGCCAACGACATCCGCTGGCTGGCAAGCGGCCCGCGCTGCGGT
TTGGGCGAAATCAAAATCCCCGAAAACGAGCCGGGTTTCGTCCATCATGCCGGGCAAAGTC
AACCCGACCCAATGCGAAGCGATGACCATGGTGTGCTGCCAAGTGTTCGGCAACGACGTT
ACCATCGGCATGGCGGGCGCGTCGGGCAATTTTCGAGCTGAACGCTCTATATGCCCGTTATC
GCCTACAACCTCTTGCAATCCATCCGCCTGTTGGGCGACGCGTGCAACAGCTTCAACGAA
CACTGCGCCGTCGGCATCGAACCCTACCGGAAAAAATCGACTATTTCTGCACCATTC
CTGATGCTGTTACTGCGTTAAACCGTAAAATCGGCTACGAAAAC

```

## 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)	dihydroliipoamide 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
atlA [NEIS2274]		DNA	18	Variable: No limits set	atlA / 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	<a href="#">Download</a>	DNA	728	Fixed: 433 bp		O. Harrison, K. Jolley	2015-06-22
adk	<a href="#">Download</a>	DNA	501	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-08
aroE	<a href="#">Download</a>	DNA	775	Fixed: 490 bp		O. Harrison, K. Jolley	2015-06-22
fumC	<a href="#">Download</a>	DNA	732	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-02
gdh	<a href="#">Download</a>	DNA	732	Fixed: 501 bp		O. Harrison, K. Jolley	2015-06-11
pdhC	<a href="#">Download</a>	DNA	747	Fixed: 480 bp		O. Harrison, K. Jolley	2015-06-25
pgm	<a href="#">Download</a>	DNA	749	Fixed: 450 bp		O. Harrison, K. Jolley	2015-06-22

**Finotyping antigens**

Locus	Download	Type	Alleles	Length	Full name/product	Curator(s)	Last updated
PorA VR1	<a href="#">Download</a>	peptide	264	Variable: No limits set	PorA variable region 1	K. Jolley	2015-06-16
PorA VR2	<a href="#">Download</a>	peptide	735	Variable: No limits set	PorA variable region 2	K. Jolley	2015-06-16
Feta VR	<a href="#">Download</a>	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	<a href="#">Download</a>	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 (hidA)	<a href="#">Download</a>	DNA	244	Variable: No limits set	D-beta-D-heptose-7-phosphate kinase	NMB0825; NMC0769; rfaE	C. Kahler	2015-05-18
NEIS0773 (hidD)	<a href="#">Download</a>	DNA	311	Variable: No limits set	ADP-D-beta-D heptose epimerase	NMB0828; NMC0773	C. Kahler	2015-05-17
NEIS2014 (gmhB)	<a href="#">Download</a>	DNA	216	Variable: No limits set	D-alpha,beta,D-Heptose 1,7 bisphosphate phosphatase	NMB2033; NMC2014	C. Kahler	2015-05-20
NEIS2055 (hidC)	<a href="#">Download</a>	DNA	206	Variable: No limits set	D-beta-D-heptose-1-phosphate adenyllyltransferase	NMB2076; NMC2055	C. Kahler	2015-05-21
NEIS2070 (gmhA)	<a href="#">Download</a>	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  
Finotyping 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

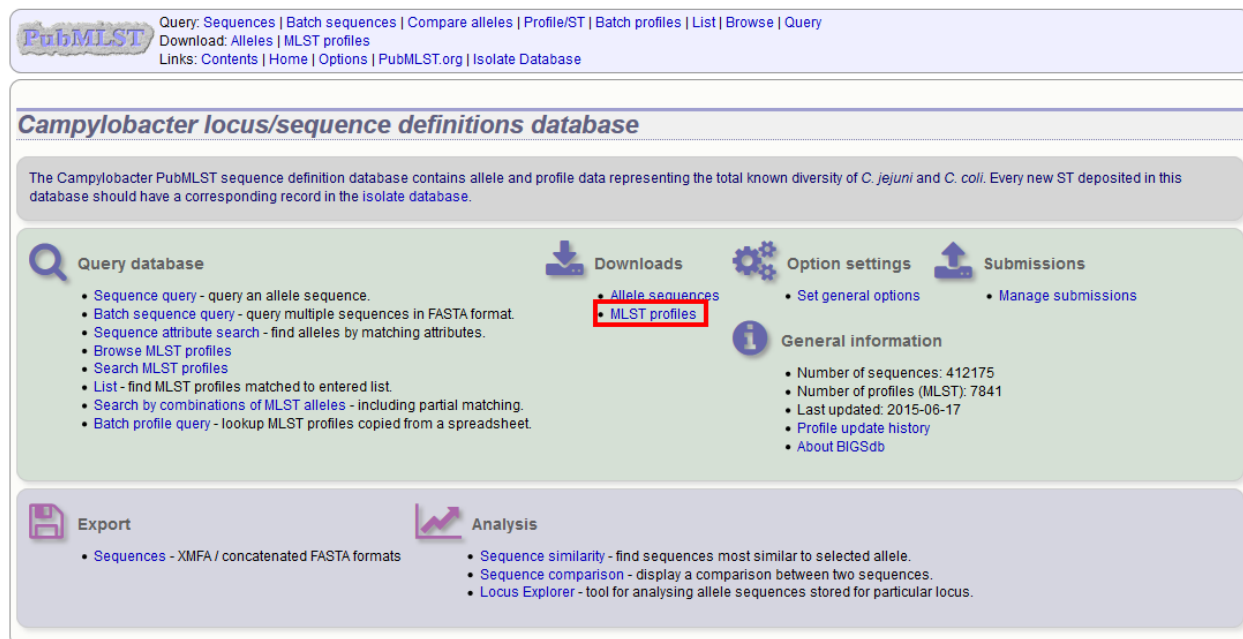
**MLST**

Locus	Download	Type	Alleles	Length	Full name/product	Curator(s)	Last updated
abcZ	<a href="#">Download</a>	DNA	728	Fixed: 433 bp		O. Harrison, K. Jolley	2015-06-22
adk	<a href="#">Download</a>	DNA	501	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-08
aroE	<a href="#">Download</a>	DNA	775	Fixed: 490 bp		O. Harrison, K. Jolley	2015-06-22
fumC	<a href="#">Download</a>	DNA	732	Fixed: 465 bp		O. Harrison, K. Jolley	2015-06-02
gdh	<a href="#">Download</a>	DNA	732	Fixed: 501 bp		O. Harrison, K. Jolley	2015-06-11
pdhC	<a href="#">Download</a>	DNA	747	Fixed: 480 bp		O. Harrison, K. Jolley	2015-06-25
pgm	<a href="#">Download</a>	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.



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

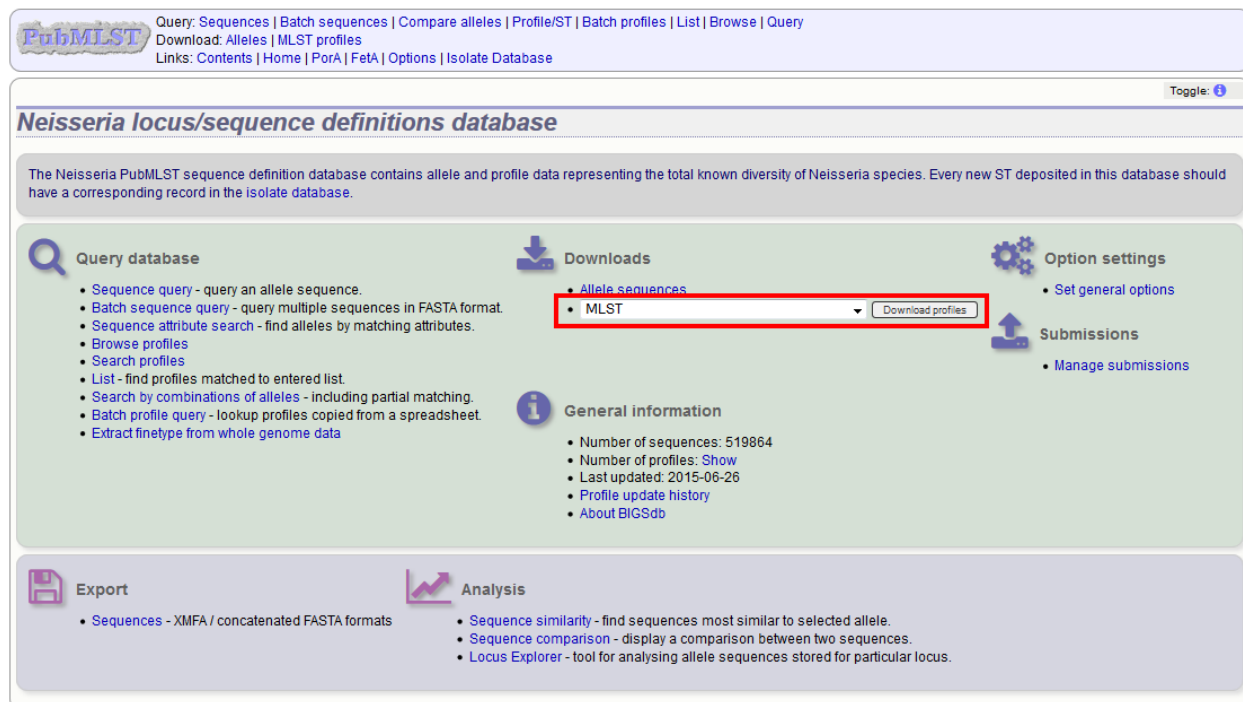
**Export**

- 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'



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

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



## CHAPTER 10

---

### Data records

---

Record pages for different types of data can be accessed following a query by clicking appropriate hyperlinks.

#### **10.1 Isolate records**

An Isolate record page displays everything known about an isolate.



[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

<b>id:</b> 18968	<b>species:</b> <i>Neisseria meningitidis</i>
<b>isolate:</b> M10 240474	<b>serogroup:</b> B
<b>strain designation:</b> B: P1.19-1,15-11: F3-9: ST-269 (cc269)	<b>ENA accession:</b> <a href="#">ERR086224</a> <a href="#">→ www.ebi.ac.uk</a>
<b>country:</b> UK	<b>sender:</b> Dorothea Hill
<b>continent:</b> Europe	<b>curator:</b> Dorothea Hill, University of Oxford, UK (E-mail: <a href="mailto:dorothea.hill@zoo.ox.ac.uk">dorothea.hill@zoo.ox.ac.uk</a> )
<b>region:</b> South East	<b>update history:</b> <a href="#">103 updates</a> <a href="#">show details</a>
<b>year:</b> 2010	<b>date entered:</b> 2012-02-15
<b>epidemiological year:</b> 07/2010-06/2011	<b>timestamp:</b> 2014-06-23
<b>disease:</b> invasive (unspecified/other)	

#### Sequence bin

<b>contigs:</b> 275	<b>N90:</b> 6405
<b>total length:</b> 2195045 bp	<b>N95:</b> 3513
<b>max length:</b> 109859 bp	<b>loci tagged:</b> 1611
<b>mean length:</b> 7982 bp	<b>detailed breakdown:</b> <a href="#">Display</a>
<b>N50:</b> 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

<b>id:</b>	18968	<b>species:</b>	Neisseria meningitidis
<b>isolate:</b>	M10 240474	<b>serogroup:</b>	B
<b>strain designation:</b>	B: P1.19-1,15-11: F3-9: ST-269 (cc269)	<b>ENA accession:</b>	ERR086224 <a href="#">→ www.ebi.ac.uk</a>
<b>country:</b>	UK	<b>sender:</b>	Dorothea Hill
<b>continent:</b>	Europe	<b>curator:</b>	Dorothea Hill, University of Oxford, UK (E-mail: <a href="mailto:dorothea.hill@zoo.ox.ac.uk">dorothea.hill@zoo.ox.ac.uk</a> )
<b>region:</b>	South East	<b>update history:</b>	<a href="#">103 updates</a> <a href="#">show details</a>
<b>year:</b>	2010	<b>date entered:</b>	2012-02-15
<b>epidemiological year:</b>	07/2010-06/2011	<b>timestamp:</b>	2014-06-23
<b>disease:</b>	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.

PubMLST

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

Update history

[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)

10.1.3 Publications

Publications (8) [show/hide](#)

- Bennett JS, Jolley KA, Sparling PF, Saunders NJ, Hart CA, Feavers IM, Maiden MC (2007). Species status of Neisseria gonorrhoeae: evolutionary and epidemiological inferences from multilocus sequence typing. *BMC Biol* **5**:35 [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 [376 isolates](#)
- Maiden MC, Bygraves JA, Feil 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**:5955-62 [78 isolates](#)
- Wang JF, Caugant DA, Li X, Hu X, Poolman JT, Crowe BA, Achtman M (1992). Clonal and antigenic analysis of serogroup A Neisseria meningitidis with particular reference to epidemiological features of epidemic meningitis in the People's Republic of China. *Infect Immun* **60**:5267-82 [47 isolates](#)

This section includes full citation for papers linked to the isolate record. Each citation has a button that will return a dataset of all isolates linked to the paper.

If there are five or more references they will be hidden by default to avoid cluttering the page too much. Click the 'Show/hide' button to display them in this case.

### 10.1.4 Sequence bin summary

**Sequence bin**

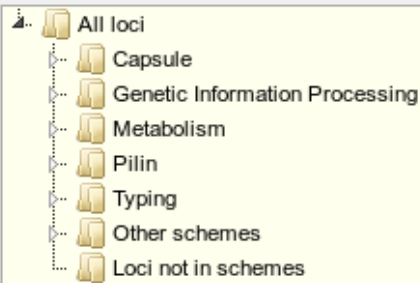
contigs:	275	N90:	6405
total length:	2195045 bp	N95:	3513
max length:	109859 bp	loci tagged:	1611
mean length:	7982 bp	detailed breakdown:	<input type="button" value="Display"/>
N50:	34308		

This section contains basic statistics describing the sequence bin. Clicking the ‘Display’ button navigates to the *sequence bin record*.

### 10.1.5 Scheme and locus data

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

**Schemes and loci**



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 <a href="#">S</a>	10 <a href="#">S</a>	15 <a href="#">S</a>	9 <a href="#">S</a>	8 <a href="#">S</a>	11 <a href="#">S</a>	9 <a href="#">S</a>	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 TTACCAACGC GTGCGCAIG AGTTGGAAGA CGGTTCGGGT GAGGCTTTGT TGAAGAAGT CAACGAATTA CAACITGAAA TCGAAGCGAA  
GGACGGCTGG AAGCTGGATG CGGCAGTCAA GCAGACTTGG GGGGAACCTG GTTTGCGCGA AAACGAAAA ATCGGCAACC TTTCGCGCG TCAGAAAAAG CGTGTGCGCT TGGCGCAGGC TTGGGTGCAG AAGCCCGAGC  
TATTGCTGCT GGACGAACCG ACCAACCATT TGGATATCGA CCGGATTATT TGGCTGGAAA ATCTGCTCAA AGCGTTTGAA GGCAGCTTGG TTGTGATTAC CCACGACCGC CGTTTTTTGG ACAATATCGC CACGCGGATT  
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](mailto: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  
contig length: 22993  
start: 13864  
end: 14607

length: 744  
orientation: forward  
complete: yes  
method: Illumina

**Sequence**

```

CACTGTACGC GGCACACGC TTCAGATTG CGGGCAGGCG GAAAACTAT TACCGTAGC CGGCGGTAA AACCGAAGT GCCGCTTAA TGGAGAAAT ATGTTAAGCG CGGCTACCT CCACCTGCAC GAAGCCTTGG GTTTGGGCC
GATGTGGTTG AAGCGGGGAG CCGTGTCTCT GCCGTCGCA ACAATTGCCG AAAGCCCGAC ACAATCCGC CGCAAAAGC AAACCGTCTT CAGCATTCCG CAGCGTCGCT CGAACAGCA TACCGGTGAG GCACGGCTCA AAACCATGAA
AGTGTGGAA ACAACCCGCG TACATACGCG CAACCCGCG CCTGAAACCG AAACGCTCTT GTCCGCGGTI TCAGACGGCA TCGCCCGCGT TCCGCGCGCT TCGGGCATAA CCAAACTTGC CGTTGTGAGC CTGTGTCCGC CGACCGAGGA
TATGTTTAC GGGCACTGT TCCACGGAAA AGCGGTGTCT CTGCTGACA ATATACTCAA AGCCGTAGGG CTGATGCGC CCTATGTCCA CAAAACCTGT TGGGTGAAAA CGCGCGCGCT CGSCACCCG ATGCCGTCTG AAGCGGCAAT
CGCAATGCG CTGGACAAA CGCGCCGGA ACTCGACGCG TGCCGCGCAC CGCGCTCTT GTTCTCGGA CAGGCGTTTG TCAACCGGA ACGGCAGGCG ATGATTGAAA CTITGTGCGC CGSCCGCCCC TTCTTCATCA TCGACCATCC
CGCCCGCTG TTGCGCCAAC CCGAACTCAA AGCCCGCGCC TGGCAGACGC TCAAGCAGCT CAAACGTGTC TTGGCGCAAG CGGCGCGCAG TTGAAGCGCG CGCACGCGG CGGTAGATC GCACTGCGT CCAATATCT GACAGAAAGC
ACAAATGAC CGATTCCGCG CAAGATTCC TAAATTCTC CTTCT

```

**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 CACTGTACGCGCGCACGGCTTCAGATTGCGGGCAGGCGGAAAACTATTACCGTAGCGCCGCGTAAACCGAAGATGCCGCTTAAATGGAGAAAAAT 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 ATGTTAAGCGCGCTACCTCCACCTGCACGAAGCCTTGGGTTTGGGCGCGATGTGGTTGAAGCGGGGAGCCGTCGTCCTGCGCTCCGCAACATTGCGCG 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 AAAGCCCGACACAAATCCGCGCGCAAAAGCAACCGTCTCAGCATTCCGCGCGTCCGTCCGACAGCATACCGGTGAGGCAAGGCTCAAAACCATGAA 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](mailto:keith.jolley@zoo.ox.ac.uk))  
date entered: 2001-02-07  
datestamp: 2009-11-11

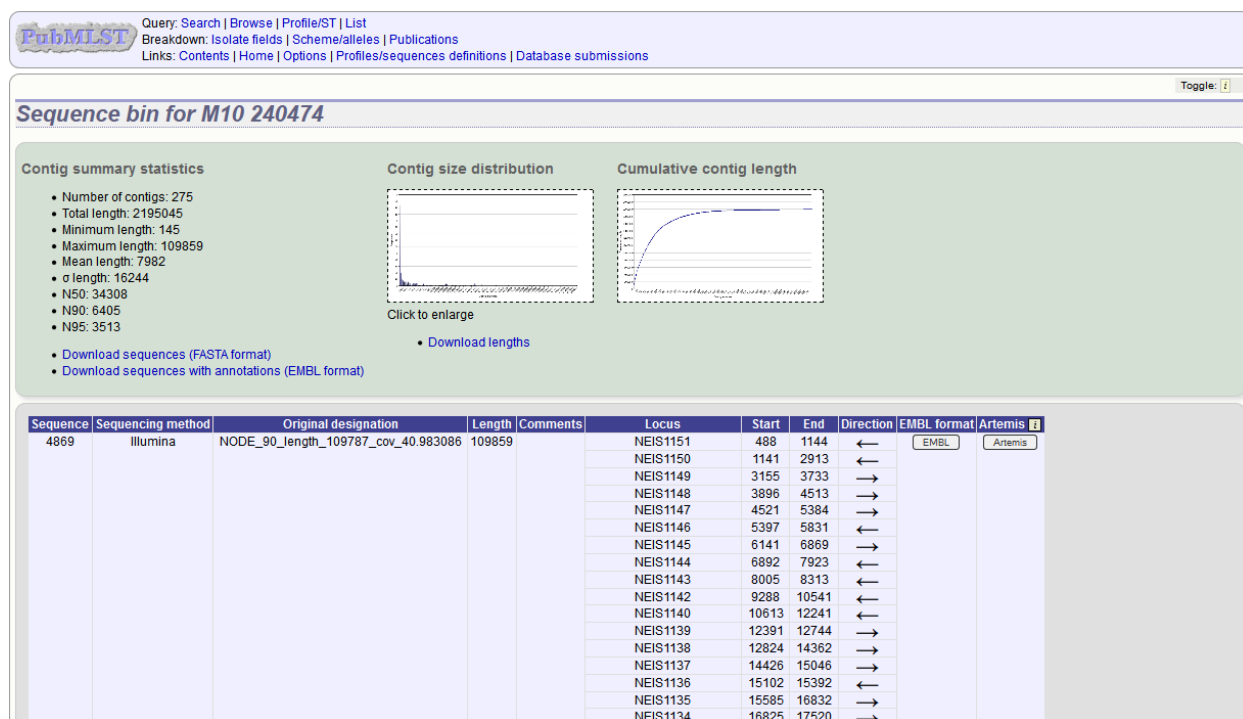
**Client database**

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:

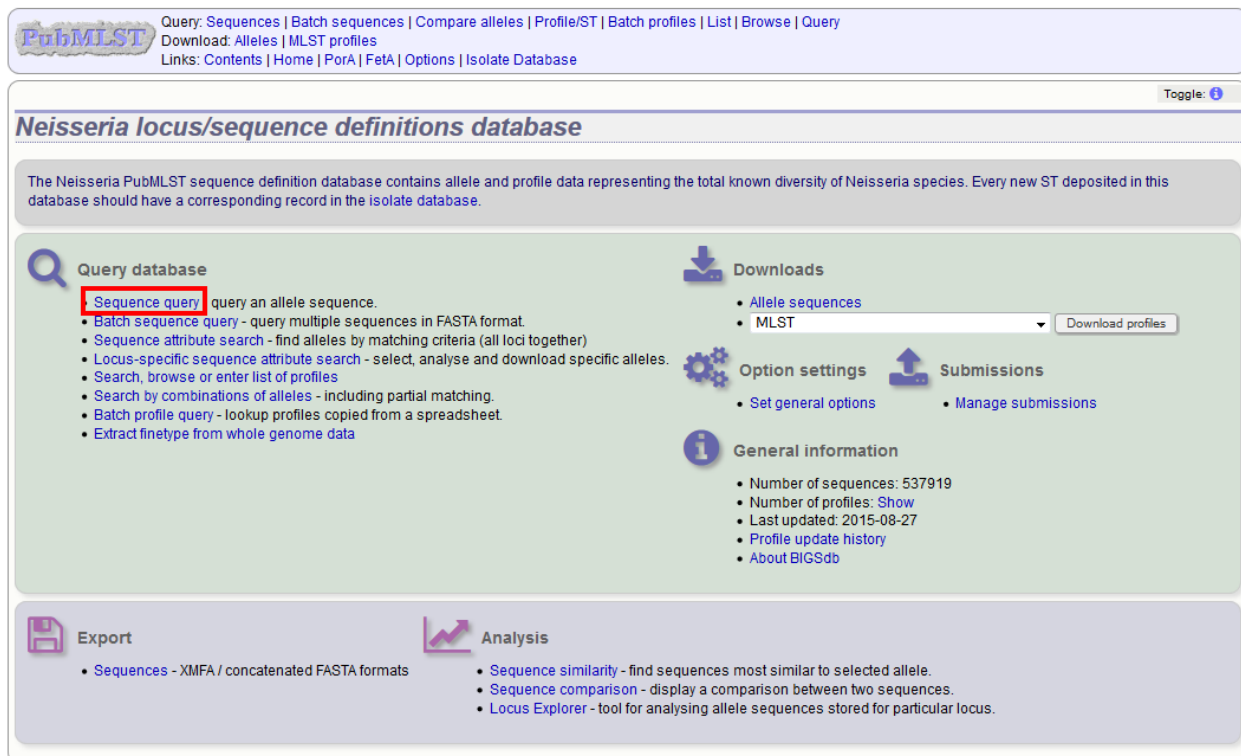
- Number of contigs
- Total length
- Minimum length
- Maximum length
- N50, N90 and N95 values
- Size distribution charts

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.

## 11.1 Querying sequences to determine allele identity

Sequence queries are performed in the sequence definition database. Click ‘Sequence query’ from the 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](#)

**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 criteria (all loci together)
- Locus-specific sequence attribute search** - select, analyse and download specific alleles.
- Search, browse or enter list of profiles**
- Search by combinations of alleles** - including partial matching.
- Batch profile query** - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data**

**Downloads**

- [Allele sequences](#)
- [MLST](#)

**Option settings**

- [Set general options](#)

**Submissions**

- [Manage submissions](#)

**General information**

- Number of sequences: 537919
- Number of profiles: [Show](#)
- Last updated: 2015-08-27
- [Profile update history](#)
- [About BIGSdb](#)

**Export**

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

**Analysis**

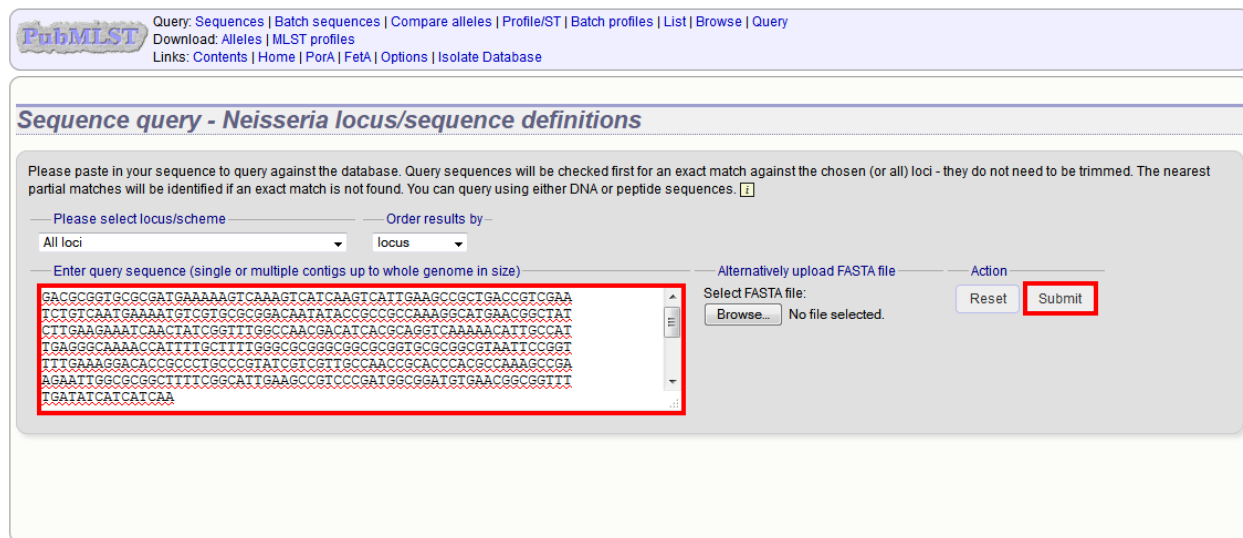
- [Sequence similarity](#) - find sequences most similar to selected allele.
- [Sequence comparison](#) - display a comparison between two sequences.
- [Locus Explorer](#) - tool for analysing allele sequences stored for particular locus.

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'.



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: All loci Order results by: locus

Enter query sequence (single or multiple contigs up to whole genome in size)

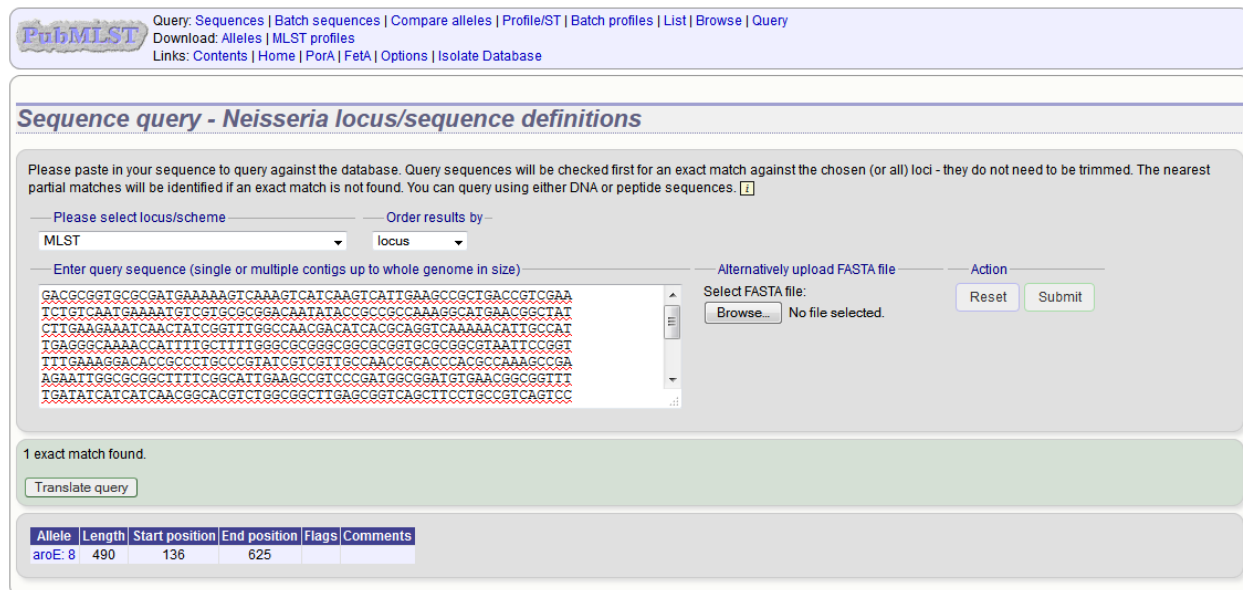
```
GACGCGGTGCGCGATGAAAAAGTCAAAGTCATCAAGTCATTGAAGCCGCTGACCGTCGAA
TCTGTCAATGAAAAATGTCGTGCGCGGACAAATATACCGCCGCAAGGSCATGAACGGCTAT
CTTGAAGAAATCAACTATCGGTITGGCCAAACGACATCACGCGAGGTCAAAAACATTGCCAT
TGAGGGCAAAACCAATTTTGCTTTTGGGCGCGGGCGCGCGGTGCGCGCGTAAATTCGGGT
TTTGAAGGACACCGCCCTGCCCCGATCGTGTGCGCAACCGCACCCACGCCAAAGCCGA
AGAATTGGCGCGGCTTTTCGGCATTGAAGCCGTCGCCGATGCGCGATGTGAACGGCGGTTT
TGATATCATCATCAA
```

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.



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)

```
GACGCGGTGCGCGATGAAAAAGTCAAAGTCATCAAGTCATTGAAGCCGCTGACCGTCGAA
TCTGTCAATGAAAAATGTCGTGCGCGGACAAATATACCGCCGCAAGGSCATGAACGGCTAT
CTTGAAGAAATCAACTATCGGTITGGCCAAACGACATCACGCGAGGTCAAAAACATTGCCAT
TGAGGGCAAAACCAATTTTGCTTTTGGGCGCGGGCGCGCGGTGCGCGCGTAAATTCGGGT
TTTGAAGGACACCGCCCTGCCCCGATCGTGTGCGCAACCGCACCCACGCCAAAGCCGA
AGAATTGGCGCGGCTTTTCGGCATTGAAGCCGTCGCCGATGCGCGATGTGAACGGCGGTTT
TGATATCATCATCAACGGCACGCTGCGCGGCTTGAGCGGTGAGCTTCCTGCCGTCAGTCC
```

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: MLST Order results by: locus

Enter query sequence (single or multiple contigs up to whole genome in size)

```
GACGCGGTGCGCGATGAAAAAGTCAAAGTCATCAGTCATTGAAGCCGCTGACCGTCGAA
TCTGTCAATGAAAAATGTCGTGCGCGGACAATATACCGCCGCAAGGSCATGAACGSCAT
CTTGAAGAAATCAACTATC8GTITIGGCCAACGACATCACGCGAGGTCAAAAACATTSCCAT
TGAGGGCAAAACCATTTTGTCTTTTGGGCGCGGGCGCGGTGCGCGCGCTAATTCCGGT
TTTGAAGGACACCGCCCTGCGCGTATCGTCGTTGCCAACCGCACCCACGCCAAAGCCGA
AGAAATTGCGCGGCTTTTCGGCATTGAAGCCGTCGCCGATGGCGGATGTGAACGGCGGTTT
TGATATCATCATCAACGGCAGCTCTGGCGGTAGAGCGGTGAGCTTCTGCCGTCAGTCC
```

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: All loci Order results by: 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
GCGGGAATGACGAAAAGCAAGCCGTAGGTCGGATACCTTGTATCCGACAAAAGCCTGCCAT
CTCAAAATAGCCGTCGGATTCCGAGATCCGACCTGCCAAAACCGGGCCGCGACGCTCCGGCC
GGCAGTTAGTACGCAATCGAACAGAACATCACAAAAAGCCCGATTCCGGATTTTCCAAAT
CGGGCTTTTTTGCGCCCGTTTTTGTATCCCGTGAAATATCCGCATGACAAAAATATAGTG
AATTACAAAAATCAGGACAAAGCGACGAAGCCGCGACAGTACAGATAGTACGGTAAGG
```

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 Querying multiple sequences to identify allele identities

You can also query multiple sequences together. These should be in FASTA format. Click ‘Batch sequence query’ from the contents page.

PubMLST Database home Contents

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

## Neisseria profile/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 criteria (all loci together)
- Locus-specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of profiles
- Search by combinations of alleles - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

### Downloads

- Allele sequences
- MLST

### Option settings

- Set general options
- Scheme options

### Submissions

- Manage submissions

### General information

- Number of sequences: 943056
- Number of profiles: [Show](#)
- Last updated: 2016-12-20
- 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 sequences (FASTA format) in to the box. Select a specific locus, scheme or 'All loci'.

PubMLST Database home Contents

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

## Batch sequence query - Neisseria profile/sequence definitions

Please paste in your sequences to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences.

Please select locus/scheme: MLST Order results by: locus

Enter query sequences (FASTA format)

```

>UIW_3
GAAGCGAAAAAATCATTGACGAAGCGGCTTGGTGCGCGACGACATCAITATCGGCATGGTCAAAGAAC
SCATCGCGCAAGACGACTGCAAAAACGGTTTCTGTTGACGGTTTCCGCGCACGCTGGCACAGGCCGA
AGCGATGGTTGAAGCAGGCGTGGATTGGATGCAGTCGTTGAAATCGACGTGCCTGACAGCGTGATTGTC
GACCGCATGAGCGGCCGCCGCTGCAITTTGGCTTCCGGCCGTACTTACCAGCTTACCTACAACCCGCCCA
AAGTTGAAGGCAAGACGACGTAACCGCGAAGATTGATTGAGCGCGACGACGACAAAGAAGAAACCGT
GAAAAACGCTTGCCTTTTACCACGACAAACCGAAGTTTGGTTGATTTTACAGCAAACTGGAAGGC
  
```

Alternatively upload FASTA file

Select FASTA file:

Browse... No file selected.

Action

Reset Submit

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

PubMLST Database home Contents

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

### Batch sequence query - Neisseria profile/sequence definitions

Please paste in your sequences to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences. [?](#)

— Please select locus/scheme — Order results by —

MLST locus

Enter query sequences (FASTA format)

```

GAAGCGAAAAATCATGACGAAGCGGCTTGGTGCSCGACGACATCATTAAGGCAATGGTCAAGAAC
GCATCGCGCAAGACGACTGCAAAAACGGTTTCTTGTITGACGGTTTCCCGCGCACATTGGCACAAAGCCGA
AGCGATGGTTGAAGCAGGCGTGGATTGGATGCACTCGTTGAAATCGAIGTGCCTGACAGCGTGAATTGTC
GACCGCATGAGCGGCGCGCGCTGCATTGGCTTCCGCGCCTACTTACCACGTTACCTACAACCGCCCA
AAGTTTAAGGCAAGACGACGTAACCGCGAAGATTGATTACGCGCGACGACGACAAAGAGAAACCGT
GAAAAAACGCCTTGCCGTTTACCACGAGCAACCGAAGTTTGGTCGATTTTACAGCAAACTGGAAGGC
GAACACGCGCCTAAATACATCAAAGTTGACGGCACCCAAGCAGTA

```

Alternatively upload FASTA file

Select FASTA file:

Browse... No file selected.

Action

Reset Submit

Sequence	Results
UIW_3	Exact match found: adk: 2
UEY_1	Partial match found: adk: 2: 2 differences found. 106 <b>T</b> → 106 <b>A</b> ; 107 <b>T</b> → 107 <b>A</b>
DSH_1	Partial match found: adk: 10: 1 difference found. 286 <b>G</b> → 286 <b>T</b>

Text format: [list](#) | [table](#)

## 11.3 Searching for specific allele definitions

There are two query pages available that allow searching for specific allele definitions. The first allows querying of all loci together by criteria that are common to all. The second is a locus-specific attribute query that can search on any extended attributes that may be defined for a locus. This locus-specific query also allows you to paste in lists of alleles for download or analysis.

### 11.3.1 General (all loci) sequence attribute search

To retrieve specific allele designations, click ‘Sequence attribute search’ on a 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 criteria (all loci together)
- Locus-specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of profiles
- Search by combinations of alleles - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

#### Downloads

- Allele sequences
- MLST

#### Option settings

- Set general options

#### Submissions

- Manage submissions

#### General information

- Number of sequences: 537919
- Number of profiles: [Show](#)
- Last updated: 2015-08-27
- [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*.

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
=
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

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


Filter query by:

Action  
Reset Submit

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

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

Filter query by:

Action  
Reset Submit

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: TTGTATACCG TTGCGAAGG TTGGGGCGAA ATTGCGGATT TATTGCGCG TTATCATCAT GTGAGCCATG AGTTGGAAAA TGGTTCGAGT GAGGCTTTGT TGAAGAGCT TAACGAATTG  
CAACTTGAAA TCGAAGCGAA GGACGCGTGG AAGCTGGATG CGSCAGTCAA GCAGACTTTG GGTGAACCTT GTTTCCAGA AACGAAAAA ATCGGCACCC TCTCGGCGG ACAGAAAAAG  
CGTGTGCCC TAGCGCAGGC TTGGGTGCAG AAGCTGATG TATTGCTGCT GGACGAACCG ACCAACCATT TGGACATTGA CGCGATTATT TGGCTGGAAA ATCTGCTTAA AGCGTTTGAA  
GGCAGCCTGG TTGTATTAC CCACGACGCG CGTTTTTTGG ACAATATGCG 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](mailto: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:

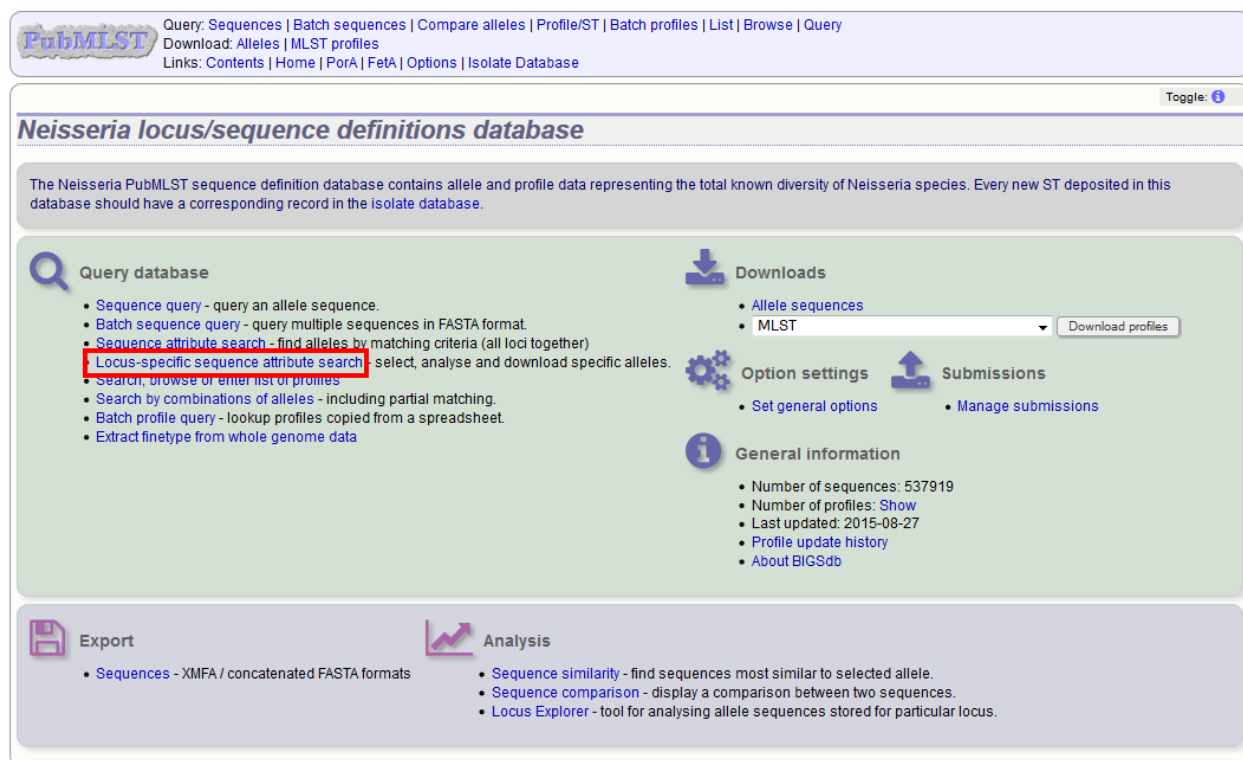
**Action**

9 records returned. Click the hyperlinks for detailed information.

locus	allele id	sequence	sequence length	comments	flags
<i>abcZ</i>	1	TTTGATACTGTTGCC ... TTGTGGAACCTCGATC	433		
<i>abcZ</i>	2	TTTGATACCGTTGCC ... TTGTGGAACCTCGATC	433		
<i>abcZ</i>	3	TTTGATACCGTTGCC ... TTGTGGAACCTGACC	433		
<i>abcZ</i>	4	TTTGATACCGTTGCC ... TTGTGGAACCTCGATC	433		
<i>abcZ</i>	5	TTTGATACCGTTGCC ... TGTGGAACCTCGATC	433		
<i>abcZ</i>	6	TTTGATACCGTTGCC ... TTGTGGAACCTCGATC	433		
<i>abcZ</i>	7	TTTGATACTGTTGCC ... TTGTGGAACCTCGATC	433		
<i>abcZ</i>	8	TTTGATACCGTTGCC ... TTGTGGAACCTGACC	433		
<i>abcZ</i>	9	TTTGATACCGTTGCC ... TTGTGGAACCTCGATC	433		

### 11.3.2 Locus-specific sequence attribute search

Some loci have *extended attribute fields*. To query these, click ‘Locus-specific sequence attribute search’ on a sequence definition database contents page.



Query: Sequences | Batch sequences | Compare alleles | Profile/ST | Batch profiles | 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 criteria (all loci together)
- Locus-specific sequence attribute search** - select, analyse and download specific alleles.
- Search, browse or enter list of profiles
- Search by combinations of alleles - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

**Downloads**

- Allele sequences
- MLST

**Option settings**

- Set general options

**Submissions**

- Manage submissions

**General information**

- Number of sequences: 537919
- Number of profiles: [Show](#)
- Last updated: 2015-08-27
- [Profile update history](#)
- [About BIGSdb](#)

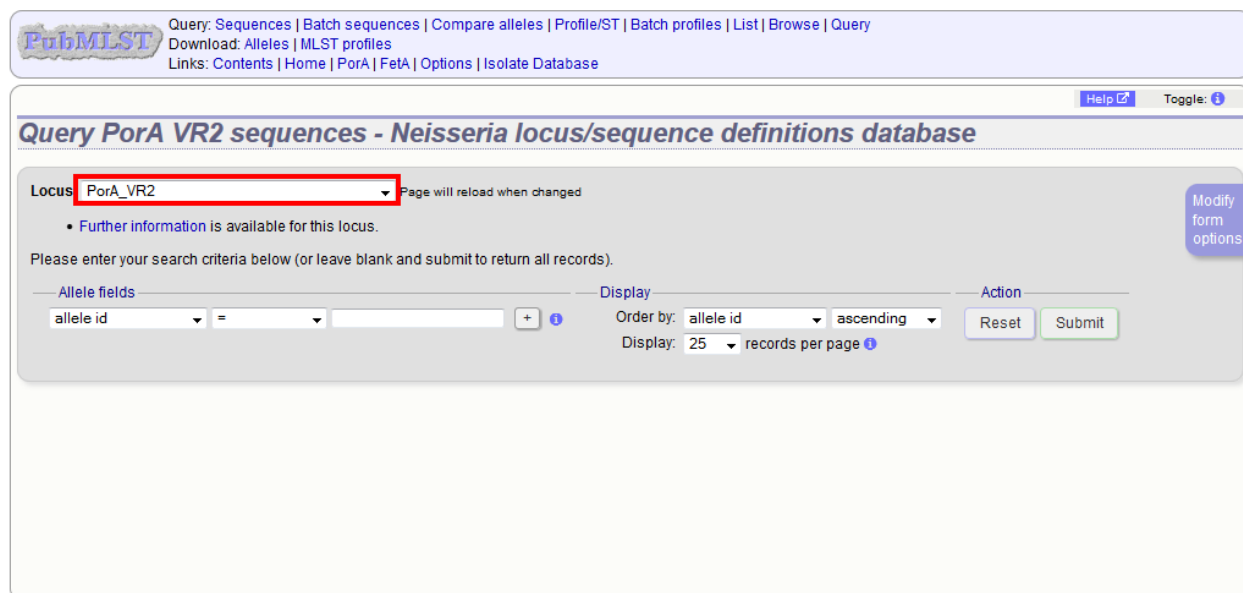
**Export**

- 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.

Pick the required locus from the dropdown box.



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

**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).

**Allele fields**

allele id =

**Display**

Order by: allele id ascending

Display: 25 records per page

**Action**

Reset Submit

Modify form options

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

Help Toggle:

### Query PorA VR2 sequences - *Neisseria* locus/sequence definitions database

Locus: PorA\_VR2 Page will reload when changed

- Further information is available for this locus.

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

Allele fields  
 family = 2 +

Display Order by: allele id ascending Action  
 Display: 25 records per page Reset Submit

84 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: Last

locus	allele id	sequence	sequence length	comments	family	variant	old name	mAb1	mAb1 reactivity	mAb2	mAb2 reactivity	flags
PorA VR2	2	HFVQQTPKSPITLVP	15		2			MN16C13F4	+	AF202	-	
PorA VR2	2-1	HFVQQPPKSPITLVP	15		2	1	2b	MN16C13F4	-	AF202	+	
PorA VR2	2-10	HFVQQAPQSQSTLVP	15		2	10						
PorA VR2	2-11	HFVLQTPQSQSTLVP	15		2	11						
PorA VR2	2-12	HFVQQIPKSPITLVP	15		2	12						
PorA VR2	2-13	YFVQQTPQSQSTLVP	15		2	13		MN16C13F4	+			
PorA VR2	2-14	HFVQQKLASKPTLVP	15		2	14	33					
PorA VR2	2-15	HFVQQKSTISKPTLVP	15		2	15	33a (33-1)					
PorA VR2	2-16	HFVQQKPTISKPTLVP	15		2	16	33b (33-2)					
PorA VR2	2-17	HFVQQQPTSEPTLVP	15		2	17	33c (33-3)					

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

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 Toggle:

### Query abcZ sequences - *Neisseria* locus/sequence definitions database

Locus: abcZ Page will reload when changed

- Further information is available for this locus.

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

Allele fields  
 allele id =  +

Display Order by: allele id ascending Action  
 Display: 25 records per page Reset Submit

Modify form options

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

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

Help ⓘ Toggle ⓘ

### Query abcZ sequences - *Neisseria* profile/sequence definitions database

Locus: abcZ Page will reload when changed

- Further information is available for this locus.

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

Allele fields

allele id =

Display

Order by: allele id ascending

Display: 25 records per page ⓘ

Modify form options

Click to add or remove additional query terms:

- Allele fields
- Allele id list box
- Filters

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

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

Help ⓘ Toggle ⓘ

### Query abcZ sequences - *Neisseria* locus/sequence definitions database

Locus: abcZ Page will reload when changed

- Further information is available for this locus.

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

Allele fields

allele id =

Allele id list

1  
2  
3  
4  
5

Display

Order by: allele id ascending

Display: 25 records per page ⓘ

Action

Reset Submit

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

• [Further information](#) is available for this locus.

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

**Allele fields**

allele id =  + [?](#)

**Allele id list**

1  
2  
3  
4  
5

**Display**

Order by: allele id ascending

Display: 25 records per page

**Action**

[Reset](#) [Submit](#)

5 records returned. Click the hyperlinks for detailed information.

locus	allele id	sequence	sequence length	comments	flags
abcZ	1	TTTGATACTGTGGC ... TTGTCGAACTCGATC	433		
abcZ	2	TTTGATAACCGTGGC ... TTGTCGAACTCGATC	433		
abcZ	3	TTTGATAACCGTGGC ... TTGTTGAACCTGACC	433		
abcZ	4	TTTGATAACCGTGGC ... TTGTCGAACTCGATC	433		
abcZ	5	TTTGATAACCGTGGC ... TCGTCGAACTCGATC	433		

**Analysis tools:**

Export: [FASTA](#) [Table](#)

Analysis: [Locus Explorer](#)

## 11.4 Browsing scheme profile definitions

If a sequence definition database has schemes defined that include a primary key field, i.e. collections of loci that together create profiles, e.g. for MLST, these can be browsed by clicking the link to ‘Search, browse or enter list of 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](#)

**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 criteria (all loci together)
- Locus specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of profiles**
- Search by combinations of alleles - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

**Downloads**

- Allele sequences
- MLST

[Download profiles](#)

**Option settings**

- Set general options

**Submissions**

- Manage submissions

**General information**

- Number of sequences: 537919
- Number of profiles: [Show](#)
- Last updated: 2015-08-27
- 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.

Leave query form fields blank (the display of these may vary depending on modification options set by the user). Choose the field to order the results by, the number of results per page to display, and click ‘Submit’.

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: ⓘ

### Search or browse profiles - *Neisseria* locus/sequence definitions

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Locus/scheme fields: ST =  + ⓘ

Display/sort options: Order by: ST ascending Display: 25 records per page ⓘ

Action: Reset Submit

Modify form options

Browsing all records.

10056 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 5 6 7 8 9 > Last

ST	adk	abcZ	aroE	fumC	gdh	pdhC	pgm	clonal complex
1	3	1	1	1	1	1	3	ST-1 complex/subgroup I/II
2	3	1	4	7	1	1	3	ST-1 complex/subgroup I/II
3	3	1	1	1	1	23	13	ST-1 complex/subgroup I/II
4	3	1	3	1	4	2	3	ST-4 complex/subgroup IV
5	1	1	2	1	3	2	3	ST-5 complex/subgroup III
6	1	1	2	1	3	2	11	ST-5 complex/subgroup III
7	1	1	2	1	3	2	19	ST-5 complex/subgroup III
8	3	2	7	2	8	5	2	ST-8 complex/Cluster A4
9	3	2	8	10	8	5	2	ST-8 complex/Cluster A4
10	3	2	4	2	8	15	2	ST-8 complex/Cluster A4
11	3	2	4	3	8	4	6	ST-11 complex/ET-37 complex
12	3	4	2	16	8	11	20	
13	10	4	15	7	8	11	4	ST-260 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 | PorA | FetA | Options | Isolate Database

Help ⓘ

### Profile information for ST-11 (MLST)

ST	adk	abcZ	aroE	fumC	gdh	pdhC	pgm	clonal complex
11	3	2	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](mailto:keith.jolley@zoo.ox.ac.uk))  
update history: [2 updates](#) [show details](#)  
date entered: 2001-02-07  
datestamp: 2013-04-27

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. [1133 isolates](#)

## 11.5 Querying scheme profile definitions

Click the link to 'Search, browse or enter list of 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

**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 criteria (all loci together)
- Locus-specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of profiles**
- Search by combinations of alleles - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

**Downloads**

- Allele sequences
- MLST

**Option settings**

- Set general options

**Submissions**

- Manage submissions

**General information**

- Number of sequences: 537919
- Number of profiles: [Show](#)
- Last updated: 2015-08-27
- Profile update history
- About BIGSdb

**Export**

- Sequences - XMTA / 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 can add search criteria by clicking the '+' button in the 'Locus/scheme fields' section. These can be combined using 'AND' or 'OR'.

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

**Search or browse profiles - Neisseria locus/sequence definitions**

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

**Locus/scheme fields**

Combine searches with: AND

date entered > 2013-02-01 +

sender (surname) = Jolley

**Display/sort options**

Order by: ST ascending

Display: 25 records per page

**Action**

Reset Submit

5051 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 5 6 7 8 9 > Last

ST	adk	abcZ	aroE	tumC	gdh	pdhC	pgm	clonal complex
5001	6	12	12	352	9	18	9	
5002	5	2	9	9	9	6	8	ST-41/44 complex/Lineage 3
5003	5	9	6	143	5	119	18	
5004	8	7	185	26	10	1	16	ST-18 complex
5005	8	7	10	19	10	351	16	ST-18 complex
5006	8	13	10	19	10	1	9	ST-18 complex
5007	8	7	10	3	9	15	20	
5008	7	12	381	91	5	21	16	
5009	8	187	10	116	10	15	20	
5010	8	7	10	17	10	1	9	ST-18 complex
5011	3	7	72	26	10	1	16	

Each field can be queried using *standard operators*.

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

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: ⓘ

### Search or browse profiles - *Neisseria* locus/sequence definitions

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

**Locus/scheme fields**      **Display/sort options**      **Action**

Combine searches with: AND ▼

date entered > 2013-02-01 + ⓘ

sender (surname) = Jolley

Order by: ST ▼ ascending ▼

Display: 25 ▼ records per page ⓘ

Reset Submit

Modify form options

5051 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 5 6 7 8 9 > Last

ST	adk	abcZ	aroE	fumC	gdh	pdhC	pgm	clonal complex
5001	6	12	12	352	9	18	9	
5002	5	2	9	9	9	6	8	ST-41/44 complex/Lineage 3
5003	5	9	6	143	5	119	18	
5004	8	7	185	26	10	1	16	ST-18 complex
5005	8	7	10	19	10	351	16	ST-18 complex
5006	8	13	10	19	10	1	9	ST-18 complex
5007	8	7	10	3	9	15	20	

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

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: ⓘ

### Search or browse profiles - *Neisseria* locus/sequence definitions

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

**Locus/scheme fields**      **Display/sort options**      **Action**

ST = + ⓘ

Order by: ST ▼ ascending ▼

Display: 25 ▼ records per page ⓘ

Reset Submit

Modify form options

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

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

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

### Search or browse profiles - Neisseria profile/sequence definitions

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Locus/scheme fields: ST =  + ?

Display/sort options: Order by: ST ascen Display: 25 records per page

**Modify form parameters**

Click to add or remove additional query terms:

- ☐ Locus/scheme field values
- ☒ Attribute values list
- ☐ Filters

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

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

### Search or browse profiles - Neisseria locus/sequence definitions

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Locus/scheme fields: ST =  + ?

**Attribute values list**

Field: ST

1  
2  
3  
4

Display/sort options: Order by: ST ascending Display: 25 records per page

Action: Reset Submit

4 records returned. Click the hyperlinks for detailed information.

ST	adk	abcZ	aroE	fumC	gdh	pdhC	pgm	clonal complex
1	3	1	1	1	1	1	3	ST-1 complex/subgroup I/II
2	3	1	4	7	1	1	3	ST-1 complex/subgroup I/II
3	3	1	1	1	1	23	13	ST-1 complex/subgroup I/II
4	3	1	3	1	4	2	3	ST-4 complex/subgroup IV

Analysis tools:

Analysis: BURST

Export: Sequences

List values will be combined with any other attributes entered in the query form allowing complex queries can be constructed.

You can also add filters to the form by again clicking the 'Modify form options' tab and selecting 'Filters'.

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

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

Help | Toggle

### Search or browse profiles - Neisseria profile/sequence definitions

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Locus/scheme fields: ST = [ ] + ⓘ

Attribute values list: Field: ST [1, 2, 3, 4] ...

Display/sort options: Order by: ST, Display: 25 records

**Modify form parameters**  
Click to add or remove additional query terms:

- ☐ Locus/scheme field values
- ☐ Attribute values list
- ☒ Filters

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

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

Toggle

### Search or browse profiles - Neisseria locus/sequence definitions

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Locus/scheme fields: ST = [ ] + ⓘ

Attribute values list: Field: ST [1, 2, 3, 4] ...

Display/sort options: Order by: ST, ascending, Display: 25 records per page

**Filters**  
Publication: [ ] ⓘ  
clonal complex: ST-4 complex/subgroup IV ⓘ

Action: Reset, Submit

1 record returned. Click the hyperlink for detailed information.

ST	adk	abcZ	aroE	fumC	gdn	pdhC	pgm	clonal complex
4	3	1	3	1	4	2	3	ST-4 complex/subgroup IV

Analysis tools:  
Export: Sequences

## 11.6 Identifying allelic profile definitions

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

Click the 'Search by combinations of alleles' link from the sequence definition contents page.

PubMLST Database home Contents

Log in Toggle: ⓘ

## Neisseria profile/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 criteria (all loci together)
- Locus-specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of profiles
- Search by combinations of alleles** - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

**Downloads**

- Allele sequences
- MLST

**Option settings**

- Set general options
- Scheme options

**Submissions**

- Manage submissions

**General information**

- Number of sequences: 943149
- Number of profiles: [Show](#)
- Last updated: 2016-12-20
- [Profile update history](#)
- [About BIGSdb](#)

**Export**

- Sequences - XMF / 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 defined in the database you should select the scheme you wish to check.

PubMLST Database home Contents

Log in Help ⓘ Toggle: ⓘ

## Search Neisseria profile/sequence definitions 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.

abcZ	adk	aroE	fumC	gdh	pdhC	pgm

Autofill profile ST:

**Options**

Search:

**Display/sort options**

Order by: ST  ascending

Display: 25  records per page ⓘ

**Action**

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

PubMLST Database home Contents

Log in Help Toggle

### Search *Neisseria* profile/sequence definitions 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.

abcZ	adk	aroE	fumC	gdh	pdhC	pgm
2	3	4	3	8	4	6

Autofill profile  
ST:

**Options**  
Search:

**Display/sort options**  
Order by: ST ascending  
Display: 25 records per page

**Action**

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

PubMLST Database home Contents

Logged in: Keith Jolley (keith) Log out Change password Help Toggle

### Search *Neisseria* profile/sequence definitions 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.

abcZ	adk	aroE	fumC	gdh	pdhC	pgm

Autofill profile  
ST: 11

**Options**  
Search:

**Display/sort options**  
Order by: ST ascending  
Display: 25 records per page

**Action**

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

PubMLST Database home Contents

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

### Search *Neisseria* profile/sequence definitions 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

abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST
2	3	4	3	8	4	6	11

**Options** **Display/sort options** **Action**

Search:  Order by:

Display:  records per page

Exact matches found (7 loci).  
1 record returned. Click the hyperlink for detailed information.

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

**Analysis tools:**  
Export:

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

PubMLST Database home Contents

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

### Search *Neisseria* profile/sequence definitions 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

abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST
2	3	4	3	8	4	6	11

**Options** **Display/sort options** **Action**

Search:  Order by:

Display:  records per page

409 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page:

ST	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	clonal complex
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
50	2	3	19	3	8	4	6	ST-11 complex/ET-37 complex
51	2	3	4	23	8	6	6	ST-11 complex/ET-37 complex
52	7	3	4	3	8	4	6	ST-11 complex/ET-37 complex
67	2	3	4	24	8	4	6	ST-11 complex/ET-37 complex
165	2	3	4	48	8	4	6	ST-11 complex/ET-37 complex
166	2	3	6	3	3	58	6	ST-11 complex/ET-37 complex
211	2	3	4	8	8	4	6	ST-11 complex/ET-37 complex
214	2	3	4	3	48	4	6	ST-11 complex/ET-37 complex
247	2	3	4	5	8	4	6	ST-11 complex/ET-37 complex
285	8	3	4	3	51	5	6	ST-11 complex/ET-37 complex

## 11.7 Batch profile queries

To lookup scheme definitions, e.g. the sequence type for multiple profiles, click ‘Batch profile query’ from the sequence definition contents page.

PubMLST Database home Contents

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

### Neisseria profile/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 criteria (all loci together)
- Locus-specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of profiles
- Search by combinations of alleles - including partial matching.
- Batch profile query** - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

**Downloads**

- Allele sequences
- MLST

**Option settings**

- Set general options
- Scheme options

**Submissions**

- Manage submissions

**General information**

- Number of sequences: 943149
- Number of profiles: [Show](#)
- Last updated: 2016-12-20
- [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 defined in the database you should select the scheme you wish to check.

PubMLST Database home Contents

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

### Batch profile query - Neisseria profile/sequence definitions

**Schemes**

Please select the scheme you would like to query:

MLST

Enter allelic profiles below in tab-delimited text format using copy and paste (for example directly from a spreadsheet). Columns can be separated by any amount of whitespace. The first column should be an isolate identifier and the remaining columns should comprise the allele numbers ([show column order](#)). Click here for [example data](#). Non-numerical characters will be stripped out of the query.

Paste in profiles

Action



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

Click submit after pasting in the data.

PubMLST Database home Contents

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

### Batch profile query - Neisseria profile/sequence definitions

**Schemes**

Please select the scheme you would like to query:

MLST

Enter allelic profiles below in tab-delimited text format using copy and paste (for example directly from a spreadsheet). Columns can be separated by any amount of whitespace. The first column should be an isolate identifier and the remaining columns should comprise the allele numbers ([show column order](#)). Click here for [example data](#). Non-numerical characters will be stripped out of the query.

Paste in profiles

isolate_6	3	6	9	5	8	6	9
isolate_7	6	6	365	55	72	6	20
isolate_8	4	3	5	4	315	21	8
isolate_9	2	273	19	3	8	4	6
isolate_10	2	7	159	92	93	6	2
isolate_11	2	5	12	143	29	285	7
isolate_12	6	7	9	56	26	18	8
isolate_13	11	5	18	5	9	24	3
isolate_14	222	231	406	12	161	2	16
isolate_15	9	4	9	17	5	3	2

Action

A results table will be displayed.

PubMLST Database home Contents

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


### Batch profile query - Neisseria profile/sequence definitions

Isolate	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST	clonal complex
isolate_1	8	3	5	4	1	3	8	290	ST-32 complex/ET-5 complex
isolate_2	2	3	4	86	8	110	118	5639	ST-11 complex/ET-37 complex
isolate_3	4	26	15	9	8	11	18	8381	ST-269 complex
isolate_4	20	5	144	35	8	130	123	942	
isolate_5	214	6	13	5	9	6	9	3312	ST-41/44 complex/Lineage 3
isolate_6	3	6	9	5	8	6	9	485	ST-41/44 complex/Lineage 3
isolate_7	6	6	365	55	72	6	20	4053	
isolate_8	4	3	5	4	315	21	8	3770	ST-32 complex/ET-5 complex
isolate_9	2	273	19	3	8	4	6	6070	ST-11 complex/ET-37 complex
isolate_10	2	7	159	92	93	6	2	10908	
isolate_11	2	5	12	143	29	285	7	4658	
isolate_12	6	7	9	56	26	18	8	10620	ST-175 complex
isolate_13	11	5	18	5	9	24	3	10202	ST-22 complex
isolate_14	222	231	406	12	161	2	16	8927	
isolate_15	9	4	9	17	5	3	2	8977	ST-103 complex

## 11.8 Investigating allele differences

### 11.8.1 Sequence similarity


To find sequences most similar to a selected allele within a sequence definition database, click ‘Sequence similarity’ on the 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](#)


---

### 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 criteria (all loci together)
- Locus-specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of MLST profiles
- Search by combinations of MLST alleles - including partial matching.
- Batch profile query - lookup MLST profiles copied from a spreadsheet.
- Extract finetype from whole genome data


**Downloads**


- Allele sequences
- MLST profiles


**Option settings**


- Set general options


**Submissions**


- Manage submissions


**General information**

- Number of sequences: 124770
- Number of profiles (MLST): 10058
- Last updated: 2015-08-19
- 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.


[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 se-

quences.

PubMLST Query: Sequences | Batch sequences | Compare alleles | Profile/ST [] | Batch profiles | Query  
Download: Alleles | MLST profiles  
Links: Contents | Home | PorA | FetA | Options | Isolate Database

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

### Find most similar alleles - *Neisseria* profile/sequence definitions

This page allows you to find the most similar sequences to a selected allele using BLAST.

Select parameters Action

Locus: abcZ  
Allele: 5  
Number of results: 10

Reset Submit

abcZ-5

Allele	% Identity	Mismatches	Gaps	Alignment	Compare
abcZ: 453	99.77	1	0	433/433	<a href="#">Compare</a>
abcZ: 405	99.77	1	0	433/433	<a href="#">Compare</a>
abcZ: 404	99.77	1	0	433/433	<a href="#">Compare</a>
abcZ: 213	99.77	1	0	433/433	<a href="#">Compare</a>
abcZ: 166	99.77	1	0	433/433	<a href="#">Compare</a>
abcZ: 114	99.77	1	0	433/433	<a href="#">Compare</a>
abcZ: 103	99.77	1	0	433/433	<a href="#">Compare</a>
abcZ: 616	99.54	2	0	433/433	<a href="#">Compare</a>
abcZ: 547	98.61	6	0	433/433	<a href="#">Compare</a>
abcZ: 342	96.54	15	0	433/433	<a href="#">Compare</a>

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: abcZ  
Allele #1: 5  
Allele #2: 453

Submit

Nucleotide differences between abcZ: 5 and abcZ: 453


Identity: 99.77 %

[Show alignment](#)

Differences: 1  
300: G → A

## 11.8.2 Sequence comparison


To directly compare two sequences click ‘Sequence comparison’ from the contents page of a sequence definition database.


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


---

### Neisseria locus/sequence definitions database


The Neisseria PubMLST sequence definition database contains allele and profile data representing the 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 criteria (all loci together)
- Locus-specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of MLST profiles
- Search by combinations of MLST alleles - including partial matching.
- Batch profile query - lookup MLST profiles copied from a spreadsheet.
- Extract finetype from whole genome data


**Downloads**


- Allele sequences
- MLST profiles


**Option settings**


- Set general options


**Submissions**


- Manage submissions


**General information**

- Number of sequences: 124770
- Number of profiles (MLST): 10058
- Last updated: 2015-08-19
- 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.


[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  
106: G → A

See also:

*Locus explorer plugin.*

## 11.9 Browsing isolate data

Isolate records can be browsed by clicking the link to ‘Search or browse database’.

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

## Neisseria PubMLST database

The Neisseria PubMLST database contains data for a collection of isolates that represent the total known diversity of Neisseria species. For 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 or browse database**
- Search by combinations of loci (profiles)
- Projects - main projects defined in database.

**Option settings**

- Set general options - including isolate table field handling.
- Set display and query options for **locus**, **schemes** or **scheme fields**.

**Submissions**

- Manage submissions

**General information**

- Isolates: 35423
- Last updated: 2015-08-27
- [Update history](#)
- [About BIGSdb](#)

**Breakdown**

- 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

Leave query form fields blank (the display of these may vary depending on modification options set by the user). Choose the field to order the results by, the number of results per page to display, and click 'Submit'.

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

Help Toggle: Field help: id Go

## Search or browse Neisseria PubMLST database

Enter search criteria or leave blank to browse all records.

Isolate provenance/phenotype fields

id = Enter value... +

Display/sort options

Order by: id ascending

Display: 25 records per page

Action

Reset Submit

Modify form options

35349 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 5 6 7 8 9 > Last

id	isolate	aliases	country	year	disease	species	serogroup	ST	MLST		Finetyping antigens		
									clonal complex	PorA VR1	PorA VR2	FetA VR	
1	A4M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex/subgroup IV	5-2	10	F1-5	
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and 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; NIBSC_3076; 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; NIBSC_2760; 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; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex/subgroup I/II	18-1	3	F5-1	
11	129E	B92; NIBSC_2828; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	A	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; NIBSC_2795; 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				

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

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

Help Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records.

Isolate provenance/phenotype fields: id = Enter value... +

Display/sort options: Order by: id ascending Display: 25 records per page

Action: Reset Submit

35349 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 5 6 7 8 9 > Last

Isolate fields										MLST		Finotyping antigens		
id	isolate	aliases	country	year	disease	species	serogroup	ST	clonal complex	PorA VR1	PorA VR2	FetA VR		
1	A4M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	4	ST-4 complex/subgroup IV	5-2	10	F1-5		
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicaemia	<i>Neisseria meningitidis</i>	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1		
3	M00242905		UK	2000	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	B	1099		19	15			
4	M1027	B43; NIBSC_3076; Z1043	USA	1937	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	4	ST-4 complex/subgroup IV					
5	M00240227		UK	2000	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	B	1100	ST-32 complex/ET-5 complex	7	16			
6	M00282207		UK	2000	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	W	1101	ST-22 complex					
7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	5	ST-5 complex/subgroup III	20	9	F3-1		
8	M00242007		UK	2000	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	B	1102	ST-18 complex		14			
9	0021/84		Czech Republic	1984	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	W	114	ST-22 complex					
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	1	ST-1 complex/subgroup I/II	18-1	3	F5-1		
11	129E	B92; NIBSC_2828; Z1092	Germany	1964	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	1	ST-1 complex/subgroup I/II	5-2	10	F3-6		
12	0090/89		Czech Republic	1989	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	B	1015	ST-32 complex/ET-5 complex	7	16			
13	139M	B99; NIBSC_2795; Z1099	Philippines	1968		<i>Neisseria meningitidis</i>	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1		
14	0120/95		Czech Republic	1995	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	X	117			14			
15	1		Germany	1999	carrier	<i>Neisseria meningitidis</i>	E	864						
16	2		Germany	1999	carrier	<i>Neisseria meningitidis</i>	B	854	ST-18 complex					
17	3		Germany	1999	carrier	<i>Neisseria meningitidis</i>	W	474	ST-474 complex					

## 11.10 Querying isolate data

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

PubMLST 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.

Query database

- Search or browse database**
- Search by combinations of loci (profiles)
- Projects - main projects defined in database.

Option settings

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

Submissions

- Manage submissions

General information

- Isolates: 35423
- Last updated: 2015-08-27
- Update history
- About BIGSdb

Breakdown

- 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** **ascending**

Display: **25** 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 <a href="#">i</a>							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:

- Provenance fields
  - Search by combination of provenance field values, e.g. country, year, sender.
- Allele designations/scheme field values
  - Search by combination of allele designations and/or scheme fields e.g. ST, clonal complex information.
- Allele designation status
  - Search by whether allele designation status is confirmed or provisional.
- Tagged sequence status
  - Search by whether tagged sequence data is available for a locus. You can also search by sequence flags.
- Attribute values list
  - Enter a list of values for any provenance field, locus, or scheme field.
- Filters
  - Various filters may be available, including
    - \* Publications
    - \* Projects
    - \* MLST profile completion status
    - \* Clonal complex
    - \* Sequence bin size
    - \* Inclusion/exclusion of *old versions*

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help ? Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter

Isolate provenance/phenotype fields

id = Enter value... + i

Display/sort options

Order by: id ascending

Display: 25 records per page i

**Modify form parameters**

Click to add or remove additional query terms:

- ☒ Provenance fields
- ☐ Allele designations/scheme field values
- ☐ Allele designation counts
- ☐ Allele designation status
- ☐ Tagged sequence counts
- ☐ Tagged sequence status
- ☐ Attribute values list
- ☐ Filters

Modify form options

If the interface is modified, a button to save options becomes available within the tab. If this is clicked, the modified form will be displayed the next time you go to the query page.

### 11.10.1 Query by allele designation/scheme field

Queries can be combined with allele designation/scheme field values.

Make sure that the allele designation/scheme field values fieldset is displayed by selecting it in the 'Modify form options' tab.

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help ? Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields

id = Enter value... + i

Allele designations/scheme fields

Display/sort options

Order by: id ascending

Display: 25 records per page i

Reset Submit

**Modify form parameters**

Click to add or remove additional query terms:

- ☒ Provenance fields
- ☒ Allele designations/scheme field values
- ☐ Allele designation counts
- ☐ Allele designation status
- ☐ Tagged sequence counts
- ☐ Tagged sequence status
- ☐ Attribute values list
- ☐ Filters

Modify form options

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

Help | Toggle: | Field help: id | Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records.

Isolate provenance/phenotype fields

Combine with: AND

country = USA

year > 1999

Allele designations/scheme fields

ST (MLST) = 11

Display/sort options

Order by: id ascending

Display: 25 records per page

Action

Reset Submit

10 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)	<i>Neisseria meningitidis</i>	W	11	ST-11 complex/ET-37 complex	5	2	F1-1
2281	M7089		USA	2000	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	W	11	ST-11 complex/ET-37 complex	5	2	F1-1
2299	M7257		USA	2000	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	W	11	ST-11 complex/ET-37 complex	5	2	F1-1
19369	M13519		USA	2005	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	W	11	ST-11 complex/ET-37 complex	5-1	10-8	F3-6
19371	M15141		USA	2006	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	C	11	ST-11 complex/ET-37 complex	5-1	10-8	
19374	M16917		USA	2007	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	NG	11	ST-11 complex/ET-37 complex	5-1	10-8	
19377	M17661		USA	2008	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	W	11	ST-11 complex/ET-37 complex	5-1	10-8	
19379	M18774		USA	2009	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	W	11	ST-11 complex/ET-37 complex	5	2	F1-94
34625	NM1495		USA	2003	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	C	11	ST-11 complex/ET-37 complex	5	2	F1-30
34640	NM313		USA	2003	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	C	11	ST-11 complex/ET-37 complex	5	2	F1-30

## 11.10.2 Query by allele designation count

Queries can be combined with counts of the total number of designations or for individual loci.

Make sure that the allele designation counts fieldset is displayed by selecting it in the ‘Modify form options’ tab.

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help | Toggle: | Field help: id | Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

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:

- Provenance fields
- Allele designations/scheme field values
- Allele designation counts
- Allele designation status
- Tagged sequence counts
- Tagged sequence status
- Attribute values list
- Filters

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

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields: id = Enter value... +

Allele designation counts: Count of total designations > 1000 +

Display/sort options: Order by: id ascending Display: 25 records per page

Action: Reset Submit

7473 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 5 6 7 8 9 > Last

Isolate fields								Seqbin size (bp)	Contigs	ST	MLST	Finotyping antigens		
id	isolate	aliases	country	year	disease	species	serogroup				clonal complex	PorA VR1	PorA VR2	FetA VR
1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	2069108	364	4	ST-4 complex/subgroup IV	5-2	10	F1-5
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicaemia	<i>Neisseria meningitidis</i>	A	2059411	359	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	2057385	199	5	ST-5 complex/subgroup III	20	9	F3-1
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	4241338	652	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
11	129E	B92; NIBSC_2828; Z1092	Germany	1964	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	2072690	272	1	ST-1 complex/subgroup I/II	5-2	10	F3-6
12	120M	B99; NIBSC_2705	Philippines	1968	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	2141744	202	4	ST-1 complex/subgroup I/II	5-2	10	F5-1

You can also search for isolates where any isolate has a particular number of designations. Use the term 'any locus' to do this.

Finally, you can search for isolates with a specific number of designations at a specific locus.

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields: id = Enter value... +

Allele designation counts: Count of abcZ > 1 +

Display/sort options: Order by: id ascending Display: 25 records per page

Action: Reset Submit

1 record returned. Click the hyperlink for detailed information.

Isolate fields								Seqbin size (bp)	Contigs	ST	MLST	Finotyping antigens		
id	isolate	aliases	country	year	disease	species	serogroup				clonal complex	PorA VR1	PorA VR2	FetA VR
34733	LNP15075		Burkina Faso	1997		<i>Neisseria meningitidis</i>		3316868	1010			20	10-1	F3-1

Analysis tools:

Breakdown: Fields Two Field Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status

Analysis: Codons Presence/Absence Genome Comparator BLAST

Export: Dataset Contigs Sequences

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

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

### 11.10.3 Query by allele designation status

Allele designations can be queried based on their status, i.e. whether they are confirmed or provisional. Queries will be combined from the values entered in all form sections.

Make sure that the allele designation status fieldset is displayed by selecting it in the ‘Modify form options’ tab.

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields: id = Enter value... +

Display/sort options: Order by: id, Display: 25 records per page

Modify form parameters: Click to add or remove additional query terms:

- Provenance fields
- Allele designations/scheme field values
- Allele designation counts
- Allele designation status**
- Tagged sequence counts
- Tagged sequence status
- Attribute values list
- Filters

Select 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 or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records.

Isolate provenance/phenotype fields: id = Enter value... +

Allele designation status: adk is provisional +

Display/sort options: Order by: id, ascending, Display: 25 records per page

Action: Reset Submit

1 record returned. Click the hyperlink for detailed information.

Isolate fields						MLST		Finotyping antigens			Loci		
id	isolate	aliases	country	year	disease	species	serogroup	ST	clonal complex	PorA VR1	PorA VR2	FetA VR	test9
19363	961-5945				Unknown	<i>Neisseria meningitidis</i>		153	ST-8 complex/Cluster A4	21	16	F5-8	

Analysis tools:

Breakdown: Fields Two Field Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status

Analysis: Codons Presence/Absence Genome Comparator BLAST

Export: Dataset Contigs Sequences

Provisional allele designations are marked within the results tables with a pink background. Any scheme field designations that depend on the allele in question, e.g. a MLST ST, will also be marked as provisional.

### 11.10.4 Query by sequence tag count

Queries can be combined with counts of the total number of tags or for individual loci.

Make sure that the tagged sequence counts fieldset is displayed by selecting it in the ‘Modify form options’ tab.

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help Toggle: Field help: id Go

**Search or browse Neisseria PubMLST database**

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields: id = Enter value... + i

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:

- Provenance fields
- Allele designations/scheme field values
- Allele designation counts
- Allele designation status
- Tagged sequence counts**
- Tagged sequence status
- Attribute values list
- Filters

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

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help Toggle: Field help: id Go

**Search or browse Neisseria PubMLST database**

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields: id = Enter value... + i

Tagged sequence counts: Count of total tags > 1000 + i

Display/sort options: Order by: id ascending Display: 25 records per page

Action: Reset Submit

7480 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 5 6 7 8 9 > Last

Isolate fields								Seqbin size (bp)	Contigs	ST	MLST		Finotyping antigens		
id	isolate	aliases	country	year	disease	species	serogroup				clonal complex	PorA VR1	PorA VR2	FetA VR	
1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	Neisseria meningitidis	A	2069108	364	4	ST-4 complex/subgroup IV	5-2	10	F1-5	
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicaemia	Neisseria meningitidis	A	2059411	359	1	ST-1 complex/subgroup I/II	5-2	10	F5-1	
7	7891	B54; NIBSC_2760; Z1054	Finland	1975	invasive (unspecified/other)	Neisseria meningitidis	A	2057385	199	5	ST-5 complex/subgroup III	20	9	F3-1	
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	Neisseria meningitidis	A	4241338	652	1	ST-1 complex/subgroup I/II	18-1	3	F5-1	
11	129E	B92; NIBSC_2828; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	A	2072690	272	1	ST-1 complex/subgroup I/II	5-2	10	F3-6	

You can also search for isolates where any isolate has a particular number of sequence tags. Use the term ‘any locus’ to do this.

Finally, you can search for isolates with a specific number of tags at a specific locus.

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help | Toggle: 1 Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields

id = Enter value... + i

Tagged sequence counts

Count of BACT000065 (rpm) = 2 + i

Display/sort options

Order by: id ascending

Display: 25 records per page

Action

Reset Submit

4564 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Page: 1 2 3 4 5 6 7 8 9 > Last

Isolate fields								Seqbin	Contigs	MLST		Finotyping antigens		
id	isolate	aliases	country	year	disease	species	serogroup	size (bp)		ST	clonal complex	PorA VR1	PorA VR2	FetA VR
1	A4/M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	2069108	364	4	ST-4 complex/subgroup IV	5-2	10	F1-5
10	6748	B73; NIBSC_2784; Z1073	Canada	1971	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	4241338	652	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
120	F4698	NIBSC_2731; Z3515	Saudi Arabia	1987	carrier	<i>Neisseria meningitidis</i>	A	4192894	638	5	ST-5 complex/subgroup III	20	9	F3-1
369	M597	NIBSC_2781; Z4323	Israel	1988	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	C	2068928	344	11	ST-11 complex/ET-37 complex	5	2-1	F5-5

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

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

### 11.10.5 Query by tagged sequence status

Sequence tags identify the region of a contig within an isolate’s sequence bin entries that correspond to a particular locus. The presence or absence of these tags can be queried as can whether or not the sequence has an a flag associated with. These flags designate specific characteristics of the sequences. Queries will be combined from the values entered in all form sections.

Make sure that the tagged sequences status fieldset is displayed by selecting it in the ‘Modify form options’ tab.

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

Help ⓘ Toggle ⓘ Field help: id Go

**Search or browse *Neisseria* PubMLST database**

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields

id = Enter value... + ⓘ

Display/sort options

Order by: id

Display: 25 records per page ⓘ

**Modify form parameters**

Click to add or remove additional query terms:

- Provenance fields
- Allele designations/scheme field values
- Allele designation counts
- Allele designation status
- Tagged sequence counts
- Tagged sequence status**
- Attribute values list
- Filters

Select 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.



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

Help Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records.

Isolate provenance/phenotype fields: id = Enter value... +

Tagged sequence status: NEIS0001 (lpxC) is flagged: internal stop codon +

Display/sort options: Order by: id ascending Display: 25 records per page

Action: Reset Submit

1 record returned. Click the hyperlink 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
2077	153	alpha153	Germany	1999	carrier	<i>Neisseria meningitidis</i>	E	60	ST-60 complex	5	2	F1-7

Analysis tools:

Breakdown: Fields Two Field Polymorphic sites Combinations Schemes/alleles Publications Sequence bin Tag status

Analysis: Codons Presence/Absence Genome Comparator BLAST

Export: Dataset Contigs Sequences

See also:

*Sequence tag flags*

### 11.10.6 Query by list of attributes

The query form can be modified with a list box in to which a list of values for a chosen attribute can be entered - this could be a list of ids, isolate names, alleles or scheme fields. This list will be combined with any other criteria or filter used on the page.

If the list box is not shown, add it by selecting it in the 'Modify form options' tab.

PubMLST Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields: id = Enter value... +

Display/sort options: Order by: id Display: 25 records per page

Modify form parameters

Click to add or remove additional query terms:

- Provenance fields
- Allele designations/scheme field values
- Allele designation counts
- Allele designation status
- Tagged sequence counts
- Tagged sequence status
- Attribute values list
- Filters

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

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

Help ? Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records.

Isolate provenance/phenotype fields  
id = Enter value... + ?

Attribute values list  
Field: id  
1  
2  
3  
4  
5

Display/sort options  
Order by: id ascending  
Display: 25 records per page

Action  
Reset Submit

5 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
1	A4M1027	B1; NIBSC_2803; Z1001	USA	1937	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	4	ST-4 complex/subgroup IV	5-2	10	F1-5
2	120M	B35; NIBSC_2822; Z1035	Pakistan	1967	meningitis and septicaemia	<i>Neisseria meningitidis</i>	A	1	ST-1 complex/subgroup I/II	5-2	10	F5-1
3	M00242905		UK	2000	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	B	1099		19	15	
4	M1027	B43; NIBSC_3076; Z1043	USA	1937	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	A	4	ST-4 complex/subgroup IV			
5	M00240227		UK	2000	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	B	1100	ST-32 complex/ET-5 complex	7	16	

Analysis tools:

Breakdown: Fields Two Field Combinations Polymorphic sites Schemes/alleles Publications Sequence bin Tag status

Analysis: BURST Codons Presence/Absence Genome Comparator BLAST

Export: Dataset Contigs Sequences

### 11.10.7 Query filters

There are various filters that can additionally be applied to queries, or the filters can be applied solely on their own so that they filter the entire database.

Make sure that the filters fieldset is displayed by selecting it in the ‘Modify form options’ tab.

Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

Help ? Toggle: Field help: id Go

### Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields  
id = Enter value... + ?

Display/sort options  
Order by: id  
Display: 25 records per page

Modify form parameters

Click to add or remove additional query terms:

- Provenance fields
- Allele designations/scheme field values
- Allele designation counts
- Allele designation status
- Tagged sequence counts
- Tagged sequence status
- Attribute values list
- Filters

The filters displayed will depend on the database and what has been defined within it. Common filters are:

- Publication - Select one or more publication that has been linked to isolate records.

- Project - Select one or more project that isolates belong to.
- Profile completion - This is commonly displayed for MLST schemes. Available options are:
  - complete - All loci of the scheme have alleles designated.
  - incomplete - One or more loci have not yet been designated.
  - partial - The scheme is incomplete, but at least one locus has an allele designated.
  - started - At least one locus has an allele designated. The scheme 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*.

## 11.11 Querying by allelic profile

If a scheme, such as MLST, has been defined for an isolate database it is possible to query the database against complete or partial allelic profiles. Even if no scheme is defined, queries can be made against all loci.

On the index page, click ‘Search by combinations of loci (profiles)’ for any defined scheme. Enter either a partial (any combination of loci) or complete profile.

The screenshot shows the Neisseria PubMLST database interface. At the top, there 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, and Database submissions. Below this is the main header 'Neisseria PubMLST database'. A descriptive paragraph follows: '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.' The main content area is divided into several sections: 'Query database' (containing 'Search or browse database', 'Search by combinations of loci (profiles)' which is redacted, and 'Projects - main projects defined in database'), 'Option settings' (containing 'Set general options - including isolate table field handling' and 'Set display and query options for locus, schemes or scheme fields'), 'Submissions' (containing 'Manage submissions'), 'General information' (containing 'Isolates: 35423', 'Last updated: 2015-08-27', 'Update history', and 'About BIGSdb'), 'Breakdown' (containing 'Single field', 'Two field', 'Unique combinations', 'Scheme and alleles', 'Publications', and 'Sequence bin'), 'Export' (containing 'Export dataset', 'Contigs', and 'Sequences - XMF / concatenated FASTA formats'), 'Analysis' (containing 'Codon usage', 'Presence/absence status of loci', 'Genome comparator', and 'BLAST'), and 'Miscellaneous' (containing '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.

PubMLST 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: Exact or nearest match

Display/sort options  
Order by: id ascending  
Display: 25 records per page

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.

PubMLST 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: Exact or nearest match

Display/sort options  
Order by: id ascending  
Display: 25 records per page

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 <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
<a href="#">41</a>	19		Germany	1999	carrier	<i>Neisseria meningitidis</i>	B	44	ST-41/44 complex/Lineage 3			
<a href="#">70</a>	38		Germany	1999	carrier	<i>Neisseria meningitidis</i>	NG	44	ST-41/44 complex/Lineage 3			
<a href="#">427</a>	NG E30	Z4692	Norway	1988	carrier	<i>Neisseria meningitidis</i>	B	44	ST-41/44 complex/Lineage 3	21	16	F1-7
<a href="#">774</a>	99 182		Canada	1999	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	B	44	ST-41/44 complex/Lineage 3			
<a href="#">792</a>	99-132		Canada	1999	invasive (unspecified/other)	<i>Neisseria meningitidis</i>	B	44	ST-41/44 complex/Lineage 3			
<a href="#">975</a>	0069/93		Czech Republic	1993	carrier	<i>Neisseria meningitidis</i>	NG	44	ST-41/44 complex/Lineage 3	22	14-4	F1-7
<a href="#">1097</a>	3532	Z7184	The Netherlands	1975		<i>Neisseria meningitidis</i>	B	44	ST-41/44 complex/Lineage 3		4	
<a href="#">1603</a>	0213/93		Czech Republic	1993	carrier	<i>Neisseria meningitidis</i>	B	44	ST-41/44 complex/Lineage 3	22	14-4	F1-19
<a href="#">1604</a>	0214/93		Czech Republic	1993	carrier	<i>Neisseria meningitidis</i>	NG	44	ST-41/44 complex/Lineage 3	22	14-4	F1-7
<a href="#">1633</a>	0244/93		Czech Republic	1993	carrier	<i>Neisseria meningitidis</i>	B	44	ST-41/44 complex/Lineage 3	22	14-4	F1-7

## 11.12 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](#) | [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.

**Query database**

- Search or browse database
- Search by combinations of loci (profiles)
- Projects - main projects defined in database.

**Option settings**

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

**Submissions**

- Manage submissions

**General information**

- Isolates: 35423
- Last updated: 2015-08-27
- Update history
- About BIGSdb

**Breakdown**

- 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:   Display:  records per page

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, Carlon 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) 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, Gray SJ, Fox AJ, Feavers IM, Maiden MC (2008) Microbiology 154:	Molecular epidemiology of meningococcal disease in England and	222 isolates

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

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
<a href="#">17517841</a>	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, 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 <i>Neisseria meningitidis</i> .	<a href="#">1670 isolates</a>
<a href="#">18815379</a>	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> .	<a href="#">1054 isolates</a>
<a href="#">15776372</a>	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.	<a href="#">822 isolates</a>
<a href="#">15528708</a>	2004	Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Cagant 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.	<a href="#">667 isolates</a>
<a href="#">17825091</a>	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.	<a href="#">576 isolates</a>
<a href="#">15537808</a>	2005	Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005) Mol Biol Evol 22: 562-9	The influence of mutation, recombination,	<a href="#">378 isolates</a>

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

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
<a href="#">17517841</a>	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, 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 <i>Neisseria meningitidis</i> .	<a href="#">1670 isolates</a>
<a href="#">18815379</a>	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> .	<a href="#">1054 isolates</a>
<a href="#">15776372</a>	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.	<a href="#">822 isolates</a>
<a href="#">15528708</a>	2004	Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Cagant 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.	<a href="#">667 isolates</a>
<a href="#">17825091</a>	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.	<a href="#">576 isolates</a>
<a href="#">15537808</a>	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> .	<a href="#">378 isolates</a>
<a href="#">15784588</a>	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.	<a href="#">271 isolates</a>
<a href="#">11101585</a>	2000	Jolley KA, Kalmusova J, Feil EJ, Gupta S, Musilek M, Kriz P, Maiden MC (2000) J Clin Microbiol 38: 4492-8	Carried meningococci in the Czech Republic: a	<a href="#">217 isolates</a>

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

## 11.13 User-configurable options

The BIGSdb user interface is configurable in a number of ways. Choices made are remembered between sessions. If the database requires you to log on, the options are associated with your user account, whereas if it is a public database, that you haven't logged in to, the options are associated with a browser cookie so they will be remembered if you connect from the same computer (using the same browser).

Most options are set by clicking the '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.

**PubMLST** 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.

**Query database**

- [Search or browse database](#)
- [Search by combinations of loci \(profiles\)](#)
- [Projects](#) - main projects defined in database.

**Option settings**

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

**Submissions**

- [Manage submissions](#)

**General information**

- Isolates: 35423
- Last updated: 2015-08-27
- [Update history](#)
- [About BIGSdb](#)

**Breakdown**

- [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](#)

### 11.13.1 General options

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

**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).

**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.

### 11.13.2 Main results table

The 'main results table' tab contains options for the display of paged results following a query.

Click the 'Main results table' header to display the tab.


[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.

### 11.13.3 Isolate record display

The 'isolate record display' tab contains options for the display of a full isolate record.

Click the 'Isolate record display' tab to display the tab.

**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.

### 11.13.4 Provenance field display

The ‘provenance field display’ tab contains checkboxes for fields to display in the main results table.

Click the ‘Provenance field display’ tab to display the tab.

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.


[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"/> datestamp
<input type="checkbox"/> sex	<input type="checkbox"/> penicillin_range	<input type="checkbox"/> ciprofloxacin	

All
None
Default
Set options

Query filters

Reset

Click the reset button to remove all user settings for this database - this includes locus and scheme field preferences.

Reset all to defaults

Some fields will be checked by default - these are defined during *database setup* (maindisplay option).

Check any fields that you wish to be displayed and then click 'Set options'. You can return to the default selection by clicking 'Default' followed by 'Set options'.

### 11.13.5 Query filters

The 'query filters' tab contains checkboxes for provenance fields and scheme completion status. Checking these results in drop-down list box filters appearing in the query page *filters fieldset*.

Click the 'Query filters' tab to display the tab.

**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"/> Finotyping 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"/> VirB T4SS profile completion
<input type="checkbox"/> amoxicillin	<input type="checkbox"/> Capsule Region A - Serogroup W profile completion	<input type="checkbox"/> beta lactamase plasmid profile completion
<input type="checkbox"/> sulphonamide	<input type="checkbox"/> Capsule Region A - Serogroup X profile completion	<input type="checkbox"/> eMLST (20 locus partial genes) profile completion
<input type="checkbox"/> ceftriaxone	<input type="checkbox"/> Capsule Region A - Serogroup Y profile completion	<input type="checkbox"/> eMLST (20 locus whole genes) profile completion
<input type="checkbox"/> ceftriaxone_range	<input type="checkbox"/> Capsule Region A - Serogroup Z profile completion	<input type="checkbox"/> rplF species profile completion
<input type="checkbox"/> chloramphenicol	<input type="checkbox"/> Capsule Region B profile completion	<input checked="" type="checkbox"/> Publications
<input type="checkbox"/> chloramphenicol_range	<input type="checkbox"/> Capsule Region C profile completion	
<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’.

### 11.13.6 Modifying locus and scheme display options

Whether or not loci, schemes or scheme fields are displayed in result tables, isolate records, or within query dropdown boxes can all be set with default options when first defined. These attributes can, however, be overridden by a user, and these selections will be remembered between sessions.

The procedure to modify these attributes is the same for locus, schemes or scheme fields, so the steps for loci will be demonstrated only.

Click the appropriate link on the isolate contents page.

The screenshot displays the Neisseria PubMLST database interface. At the top, there is a navigation bar with links for Query, Breakdown, and Links. Below this, a header section identifies the database as 'Neisseria PubMLST database' and provides a brief description of its contents. The main area is divided into several functional sections: 'Query database' (with links for searching and browsing), 'Option settings' (where 'locus, schemes or scheme fields' is highlighted in a red box), 'Submissions' (for managing submissions), 'General information' (showing isolate counts and update history), 'Breakdown' (for viewing data by field or scheme), 'Export' (for downloading datasets), 'Analysis' (for codon usage and BLAST), and 'Miscellaneous' (for field descriptions).

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](#)

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:  =

Display: Order by:     
 Display:  records per page

Filter query by:

1 record returned. Click the hyperlink for detailed information.

Customize:

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:  =

Display: Order by:     
 Display:  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:

7 records returned. Click the hyperlinks for detailed information.

Customize:

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*
<a href="#">abcZ</a>	DNA	integer	433	false	true	1	1176340	allele only	false	true	true
<a href="#">adk</a>	DNA	integer	465	false	true	1	991951	allele only	false	true	true
<a href="#">aroE</a>	DNA	integer	490	false	true	2	2079469	allele only	false	true	true
<a href="#">fumC</a>	DNA	integer	465	false	true	1	1592943	allele only	false	true	true
<a href="#">gdh</a>	DNA	integer	501	false	true	1	1514394	allele only	false	true	true
<a href="#">pdhC</a>	DNA	integer	480	false	true	1	1453970	allele only	false	true	true
<a href="#">pgm</a>	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).



## CHAPTER 12

### User projects

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

**Note:** User projects can be enabled by an administrator by setting ‘user\_projects=”yes”’ in the *config.xml* file for the database.

To create a new project, go to the isolate database contents page and click ‘Your projects’.

The screenshot shows the Neisseria PubMLST database homepage. At the top, there's a navigation bar with 'PubMLST', 'Database home', and 'Contents'. Below this, a login status bar indicates 'Logged in: Keith Jolley (keith)' with links for 'Log out' and 'Change password'. The main heading is 'Neisseria PubMLST database'. A descriptive paragraph follows: '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.'

The interface is divided into several functional sections:

- Query database**: Includes links for 'Search or browse database' and 'Search by combinations of loci (profiles)'.
- Projects**: Contains two links: 'Main public projects' and 'Your projects', which is highlighted with a red rectangle.
- Option settings**: Includes links for 'Set general options', 'Set display and query options', and 'View/modify client software permissions'.
- Submissions**: Includes a link for 'Manage submissions'.
- General information**: Displays statistics such as 'Isolates: 43053', 'Last updated: 2017-06-19', and links for 'Defined field values', 'Update history', and 'About BIGSdb'.
- Breakdown**: Includes links for 'Single field', 'Two field', 'Unique combinations', 'Scheme and alleles', 'Publications', and 'Sequence bin'.
- Export**: Includes links for 'Export dataset', 'Contigs', and 'Sequences - XMFA / concatenated FASTA formats'.
- Analysis**: Includes links for 'Codon usage', 'Presence/absence status of loci', 'Genome comparator', and 'BLAST'.
- Miscellaneous**: Includes a link for 'Description of database fields'.

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

PubMLST Database home Contents

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

## User projects

### New private projects

Projects allow you to group isolates so that you can analyse them easily together.

Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).

**New project** Action

Name:

Description:

### Existing projects

You do not own or are a member of any projects.

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

## 12.1 Adding isolates to a user project following a query

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

PubMLST Database home Contents

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

## Search or browse *Neisseria* PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

**Isolate provenance/phenotype fields** Allele designations/scheme fields

Combine with: **AND**

country = UK

serogroup = W

year = 2016

ST (MLST) = 11

**Display/sort options** Action

Order by: id

Display: 25 records per page

163 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

**Your projects**

UK group W ST-11 2016

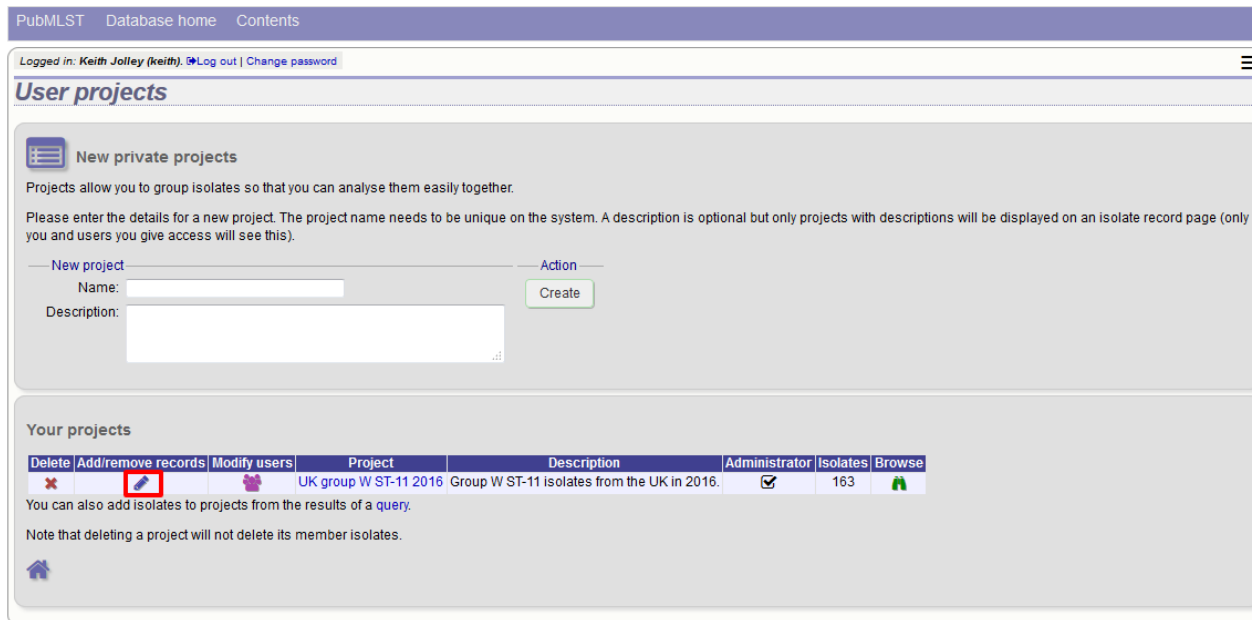
Page: 1 2 3 4 5 6 7 > Last

Isolate fields									MLST											Finotyping antigens		
id	isolate	aliases	country	year	disease	source	species	serogroup	Seqbin size (bp)	Contigs	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST	clonal complex	PorA VR1	PorA VR2	FetA VR
42451	M16 240077		UK	2016		sputum	<i>Neisseria meningitidis</i>	W	2111799	257	2	3	4	3	8	4	6	11	ST-11 complex/ET-37 complex	5	2	F1-1
42475	M16 240003		UK	2016	invasive (unspecified/other)		<i>Neisseria meningitidis</i>	W	2098168	222	2	3	4	3	8	4	6	11	ST-11 complex/ET-37 complex	5	2	F1-1
42478	M16 240008		UK	2016	invasive (unspecified/other)		<i>Neisseria meningitidis</i>	W	2108815	265	2	3	4	3	8	4	6	11	ST-11 complex/ET-37 complex	5	2	F1-1
42480	M16 240010		UK	2016	invasive (unspecified/other)		<i>Neisseria meningitidis</i>	W	2117414	338	2	3	4	3	8	4	6	11	ST-11 complex/ET-37 complex	5	2	F1-1
42481	M16		UK	2016	invasive		<i>Neisseria</i>	W	2106704	269	2	3	4	3	8	4	6	11	ST-11	5	2	F1-146

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

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

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



PubMLST Database home Contents

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

### User projects

**New private projects**

Projects allow you to group isolates so that you can analyse them easily together.

Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).

New project Action

Name:

Description:

Create

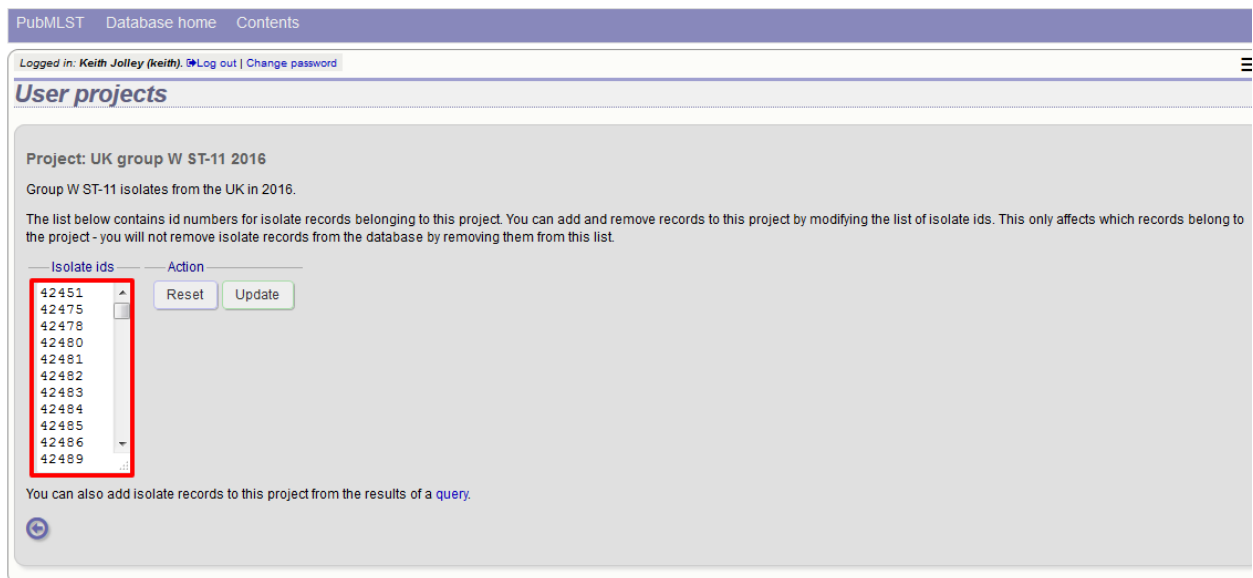
**Your projects**

Delete	Add/remove records	Modify users	Project	Description	Administrator	Isolates	Browse
			UK group W ST-11 2016	Group W ST-11 isolates from the UK in 2016.	<input checked="" type="checkbox"/>	163	

You can also add isolates to projects from the results of a [query](#).

Note that deleting a project will not delete its member isolates.

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



PubMLST Database home Contents

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

### User projects

**Project: UK group W ST-11 2016**

Group W ST-11 isolates from the UK in 2016.

The list below contains id numbers for isolate records belonging to this project. You can add and remove records to this project by modifying the list of isolate ids. This only affects which records belong to the project - you will not remove isolate records from the database by removing them from this list.

Isolate ids Action

42451  
42475  
42478  
42480  
42481  
42482  
42483  
42484  
42485  
42486  
42489

Reset Update

You can also add isolate records to this project from the results of a [query](#).

## 12.3 Accessing project isolates

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

PubMLST Database home Contents

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

### User projects

**New private projects**

Projects allow you to group isolates so that you can analyse them easily together.

Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).

New project:  Name:  Description:

Action:

**Your projects**

Delete	Add/remove records	Modify users	Project	Description	Administrator	Isolates	Browse
			UK group W ST-11 2016	Group W ST-11 isolates from the UK in 2016.	<input checked="" type="checkbox"/>	163	

You can also add isolates to projects from the results of a query.

Note that deleting a project will not delete its member isolates.

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

PubMLST Database home Contents

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

### Search or browse Neisseria PubMLST database

Enter search criteria or leave blank to browse all records. Modify form parameters to filter or enter a list of values.

Isolate provenance/phenotype fields: id = Enter value... +

Filters:

- Publication: 1 selected
- Project: 1 selected
- MLST profiles: ☒ Check all ☒ Uncheck all
- Clonal complex (MLST): ☐ Irish Meningococcus Genome Library
- Bexsero OMV peptides...: ☐ MRF Meningococcus Genome Library
- Ribosomal MLST profiles: ☐ NI\_EXCL\_MRF
- Sequence bin: ☐ penA sequence analysis
- ☐ rpoB sequence analysis
- ☐ Include old record versions
- Scottish MRF Genome Library
- ☒ UK group W ST-11 2016
- ☐ UK & Ireland MGL 2010/11-2013/14

Display/sort options: Order by: id ascending Display: 25 records per page

163 records returned (1 - 25 displayed). Click the hyperlinks for detailed information.

Your projects: Select project... Add these records

Page: 1 2 3 4 5 6 7 Last

Isolate fields								Seqbin	Contigs	MLST										Finotyping antigens		
id	isolate	aliases	country	year	disease	source	species	serogroup	size (bp)	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST	clonal complex	PorA VR1	PorA VR2	FetA VR	
42451	M16		UK	2016		sputum	Neisseria meningitidis	W	2111799	257	2	3	4	3	8	4	6	11	ST-11 complex/ET-37	5	2	F1-1

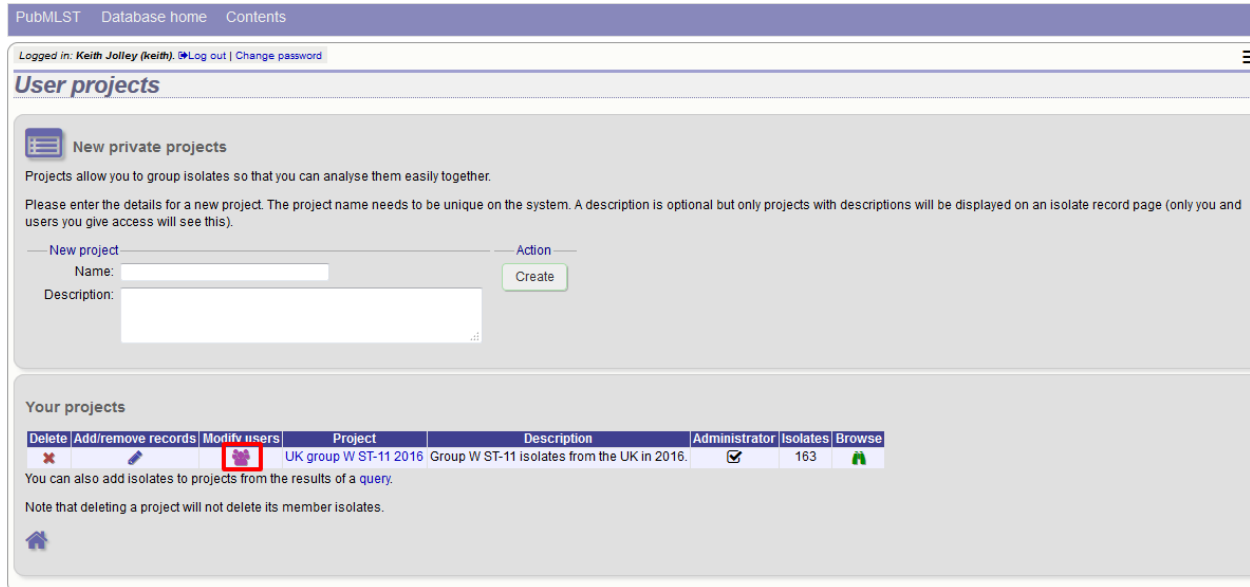
Please note that you will only see your project in the filter list if you are logged in. As a private project, only you will see it.



## 12.4 Allowing other users to share your project

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

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



PubMLST Database home Contents

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

### User projects

**New private projects**

Projects allow you to group isolates so that you can analyse them easily together.

Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).

New project

Name:

Description:

Action

Create

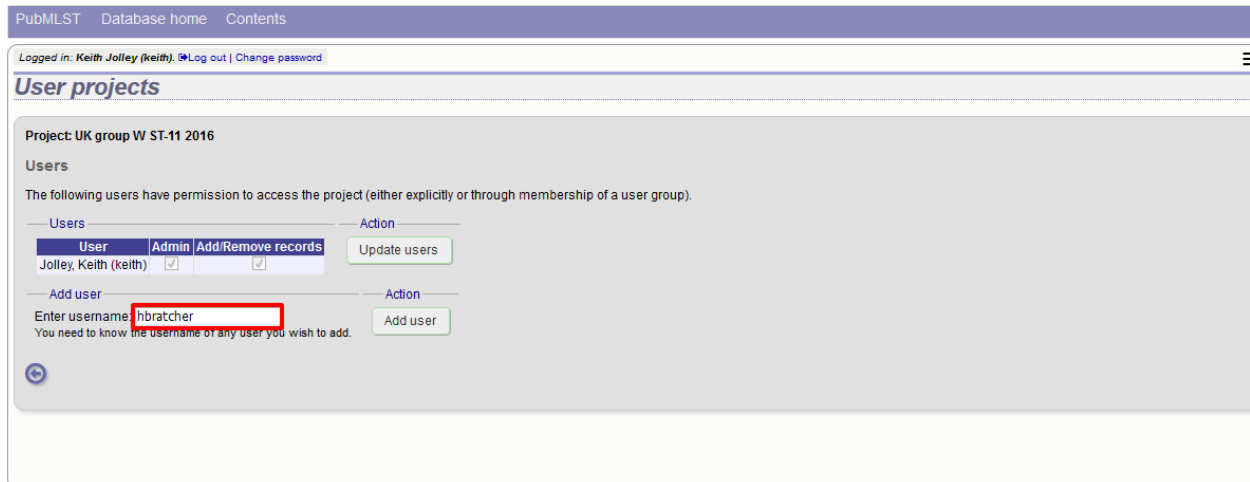
**Your projects**

Delete	Add/remove records	Modify users	Project	Description	Administrator	Isolates	Browse
			UK group W ST-11 2016	Group W ST-11 isolates from the UK in 2016.		163	

You can also add isolates to projects from the results of a [query](#).

Note that deleting a project will not delete its member isolates.

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



PubMLST Database home Contents

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

### User projects

**Project: UK group W ST-11 2016**

**Users**

The following users have permission to access the project (either explicitly or through membership of a user group).

Users

User	Admin	Add/Remove records
Jolley, Keith (keith)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Action

Update users

**Add user**

Enter username:

You need to know the username of any user you wish to add.

Action

Add user

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

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

PubMLST Database home Contents

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

## User projects

Project: UK group W ST-11 2016

Users

The following users have permission to access the project (either explicitly or through membership of a user group).

Remove	User	Admin/Add/Remove records	Action
	Bratcher, Holly (hbratcher)	<input type="checkbox"/>	<input type="button" value="Update users"/>
	Jolley, Keith (keith)	<input checked="" type="checkbox"/>	

Add user

Enter username: hbratcher

You need to know the username of any user you wish to add.

## 12.5 Deleting a user project

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

PubMLST Database home Contents

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

## User projects

New private projects

Projects allow you to group isolates so that you can analyse them easily together.

Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).

New project

Name:

Description:

Your projects

Delete	Add/remove records	Modify users	Project	Description	Administrator	Isolates	Browse
			UK group W ST-11 2016	Group W ST-11 isolates from the UK in 2016.	<input checked="" type="checkbox"/>	163	

You can also add isolates to projects from the results of a [query](#).

Note that deleting a project will not delete its member isolates.

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

PubMLST Database home Contents

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

## User projects

This project contains 163 isolates. Please confirm that you wish to remove the project (the isolates in the project will not be deleted).

**Delete project**

### New private projects

Projects allow you to group isolates so that you can analyse them easily together.

Please enter the details for a new project. The project name needs to be unique on the system. A description is optional but only projects with descriptions will be displayed on an isolate record page (only you and users you give access will see this).

New project Action

Name:

Description:

Create

### Your projects

Delete	Add/remove records	Modify users	Project	Description	Administrator	Isolates	Browse
			UK group W ST-11 2016	Group W ST-11 isolates from the UK in 2016.	<input checked="" type="checkbox"/>	163	

You can also add isolates to projects from the results of a [query](#).

Note that deleting a project will not delete its member isolates.

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

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



# CHAPTER 13

## Private records

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

### 13.1 Uploading private records

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

The screenshot displays the Neisseria PubMLST database homepage. At the top, a navigation bar includes 'PubMLST', 'Database home', and 'Contents'. Below this, a status bar indicates the user is logged in as 'Keith Jolley (keith)' with links for 'Log out' and 'Change password'. The main heading is 'Neisseria PubMLST database'. A descriptive paragraph states: '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.'

The interface is organized into several functional blocks:

- Query database**: Search or browse database, Search by combinations of loci (profiles).
- Projects**: Main public projects, Your projects.
- Option settings**: Set general options - including isolate table field handling, Set display and query options for locus, schemes or scheme fields, View/modify client software permissions.
- Submissions**: Manage submissions.
- Private data**: Upload/manage records (this link is highlighted with a red box).
- General information**: Isolates: 43172, Last updated: 2017-06-23, Defined field values, 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.

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

PubMLST Database home Contents

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

## Private records

**Limits**

Accounts have a quota for the number of private records that they can upload. Uploading of private data to some registered projects may not count against your quota.

Records (total): 0  
Records (quota): 0  
Quota: 100  
You can upload: 100

**Upload**

[Upload private isolate records](#)

**Curate**

- [Update private records](#) Curator's interface

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

PubMLST Database home Contents

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

## Batch insert isolates

**Private data upload**

These isolates will count against your quota of private data.

Quota available: 100

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 \(xlsx format\)](#)

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

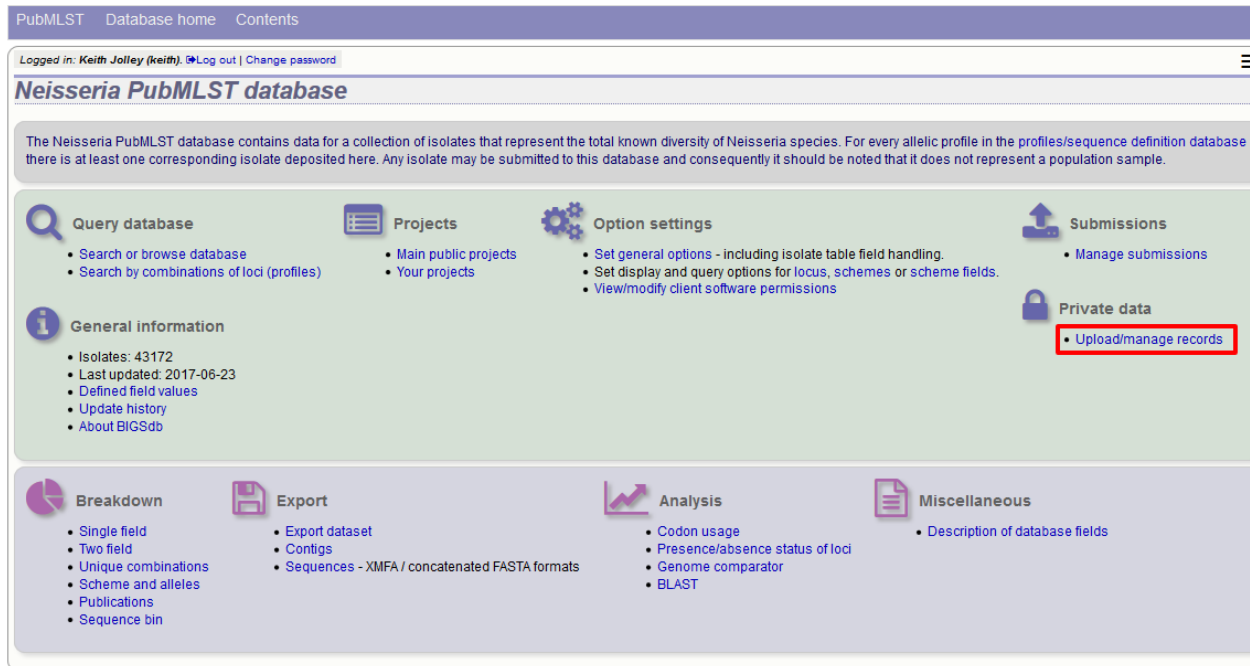
Action

[Reset](#) [Submit](#)

See [batch adding isolate records](#) for details of the upload process.

## 13.2 Modifying private records

Click the ‘Upload/manage records’ link on the contents page.



PubMLST Database home Contents

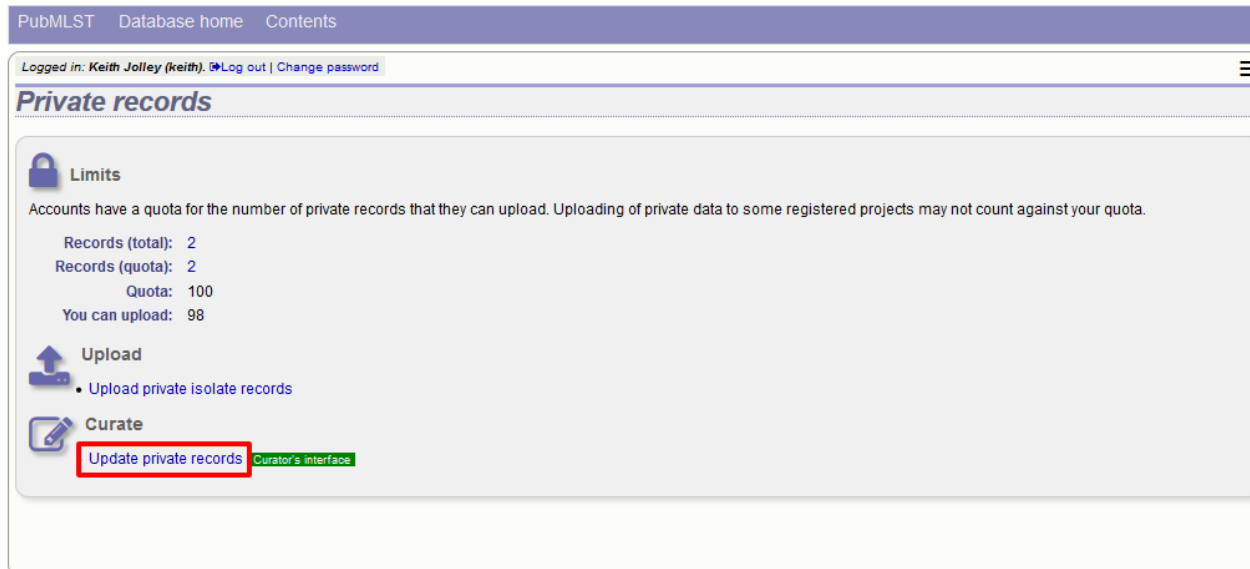
Logged in: Keith Jolley (keith). Log out | Change password

### 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 or browse database
  - Search by combinations of loci (profiles)
- Projects**
  - Main public projects
  - Your projects
- Option settings**
  - Set general options - including isolate table field handling.
  - Set display and query options for locus, schemes or scheme fields.
  - View/modify client software permissions
- Submissions**
  - Manage submissions
- General information**
  - Isolates: 43172
  - Last updated: 2017-06-23
  - Defined field values
  - Update history
  - About BIGSdb
- Private data**
  - Upload/manage records
- 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

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



PubMLST Database home Contents

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

### Private records

**Limits**

Accounts have a quota for the number of private records that they can upload. Uploading of private data to some registered projects may not count against your quota.

Records (total):	2
Records (quota):	2
Quota:	100
You can upload:	98

**Upload**

- Upload private isolate records

**Curate**

- Update private records Curator's interface

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

## 13.3 Sharing access to private records

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

See *user projects* for more details.



## 14.1 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

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**

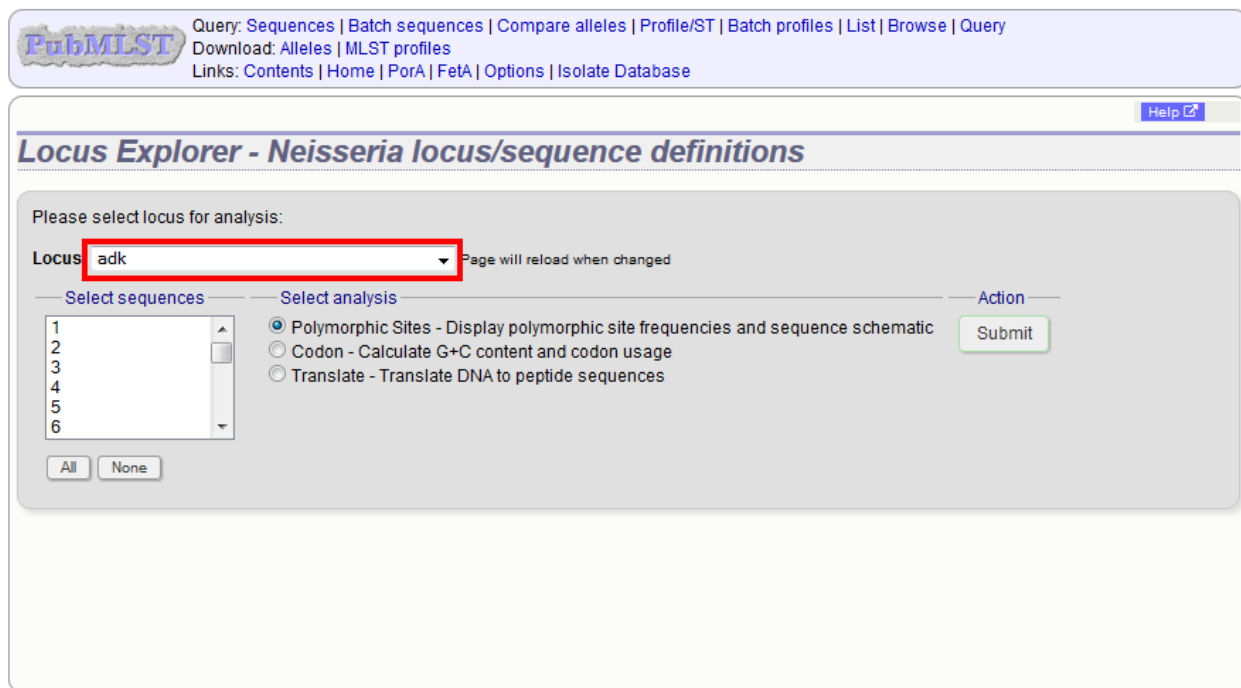
- Sequences - XIMFA / concatenated FASTA formats

**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.

### 14.1.1 Polymorphic site analysis

Select the locus you would like to analyse in the Locus dropdown box. The page will reload.



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**

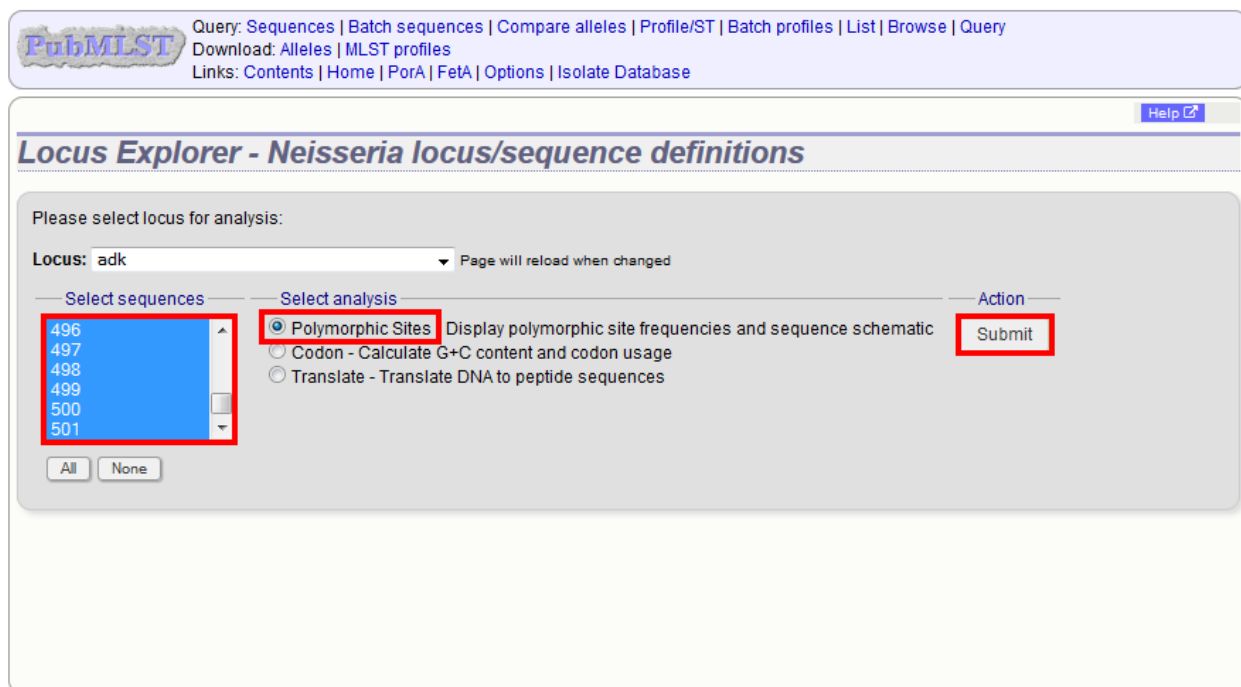
- 1
- 2
- 3
- 4
- 5
- 6

**Select analysis**

- ☒ Polymorphic Sites - Display polymorphic site frequencies and sequence schematic
- ☐ Codon - Calculate G+C content and codon usage
- ☐ Translate - Translate DNA to peptide sequences

**Action**

Select the alleles that you would like to include in the analysis. Variable length loci are limited to 2000 sequences or fewer since these need to be aligned. Select 'Polymorphic Sites' in the Analysis selection and click 'Submit'.



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**

- 496
- 497
- 498
- 499
- 500
- 501


**Select analysis**

- ☒ Polymorphic Sites - Display polymorphic site frequencies and sequence schematic
- ☐ Codon - Calculate G+C content and codon usage
- ☐ Translate - Translate DNA to peptide sequences

**Action**

If an alignment is necessary, the job will be submitted to the job queue and the analysis performed. If no alignment is





[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](#)

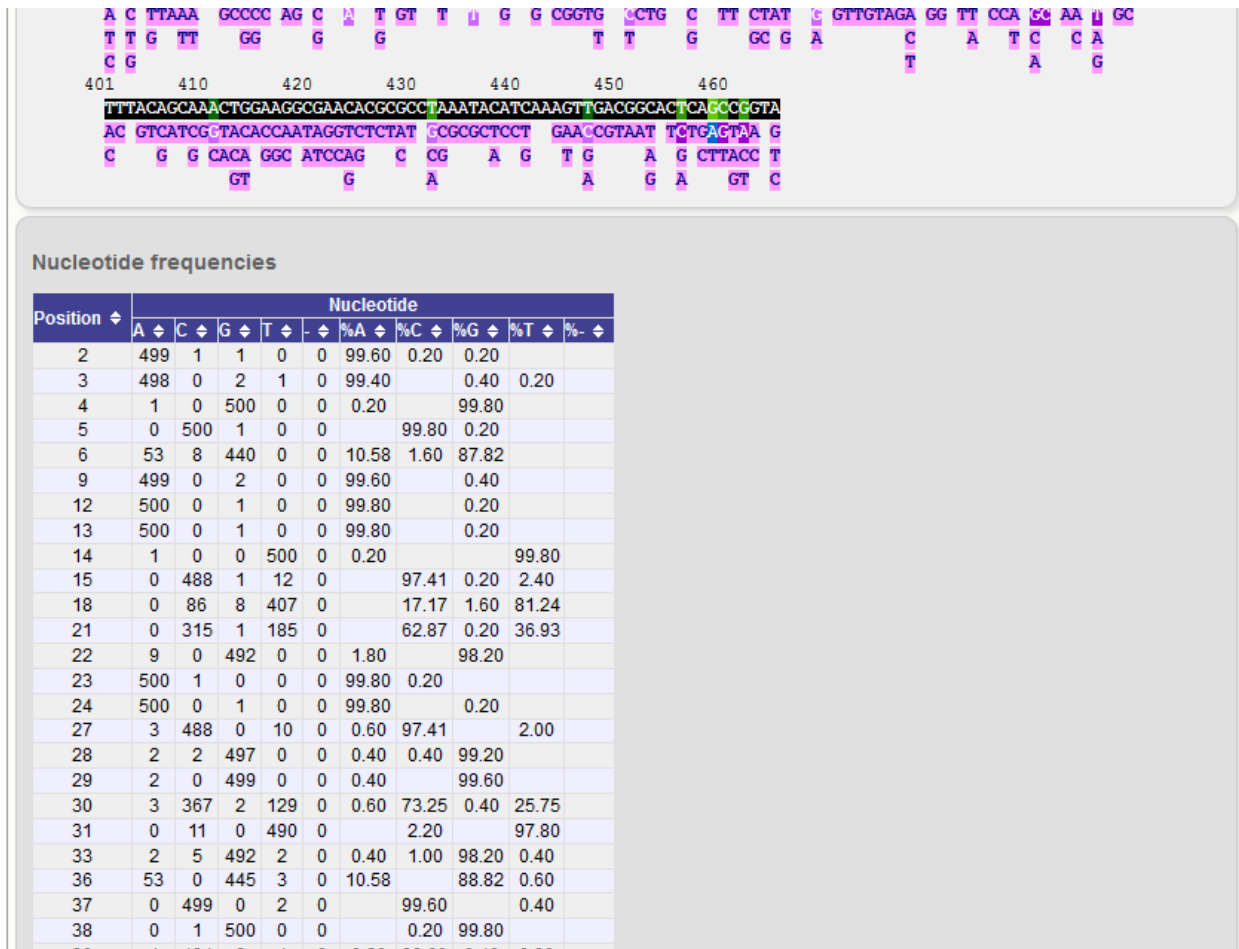
## Site Explorer

**adk position 51**

501 alleles included in analysis.

Base	Number of alleles	Percentage of alleles	MLST profiles
T	401	80.04	11130 / 11407 (97.57%)
C	98	19.56	275 / 11407 (2.41%)
A	1 (adk-351)	0.20	1 / 11407 (0.01%)
G	1 (adk-413)	0.20	1 / 11407 (0.01%)

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



See also:

- *Investigating allele differences.*
- *Polymorphism analysis following isolate query.*

### 14.1.2 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	0.606	46.786	3333

### 14.1.3 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:  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

An aligned amino acid sequence will be displayed.

```

459_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 EAKKIIDEGLVRDDIIIGMVKERIAQDDCKNGFLFDGFPRTLAQAEAMVEAGVDLDAVVEIDVPDSVIVDRMSGRVRVHLASGRTHVVTYNPPKVEGKDD
          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.

## 14.2 Field breakdown

The field breakdown plugin for isolate databases displays the frequency of each value for fields, alleles and schemes.

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.

15	1		Republic	(unspecified/other)	meningitidis									
			Germany	1999	carrier	Neisseria meningitidis	E	0	0	864				
16	2		Germany	1999	carrier	Neisseria meningitidis	B	0	0	854	ST-18 complex			
17	3		Germany	1999	carrier	Neisseria meningitidis	W	0	0	174	ST-174 complex			
18	4		Germany	1999	carrier	Neisseria meningitidis	B	0	0	19	ST-18 complex			
19	S3131	B213; NIBSC_2813; Z1213	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	A	2061338	173	4	ST-4 complex	7	13-1	F1-5
20	5		Germany	1999	carrier	Neisseria meningitidis	NG	0	0	198	ST-198 complex			
21	6		Germany	1999	carrier	Neisseria meningitidis	NG	0	0	198	ST-198 complex			
22	7		Germany	1999	carrier	Neisseria meningitidis	E	0	0	60	ST-60 complex			
23	8		Germany	1999	carrier	Neisseria meningitidis	B	0	0	32	ST-32 complex			
24	S4355	B227; NIBSC_2806; Z1227	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	A	2071992	198	5	ST-5 complex	5-1	9	F3-1
25	9		Germany	1999	carrier	Neisseria meningitidis	B	0	0	930	ST-334 complex			

**Analysis tools:**

Breakdown: **Fields** | Two Field | Combinations | Polymorphic sites | Publications | Sequence bin | Tag status

Analysis: BURST | Codons | Presence/Absence | Genome Comparator | BLAST | rMLST species id

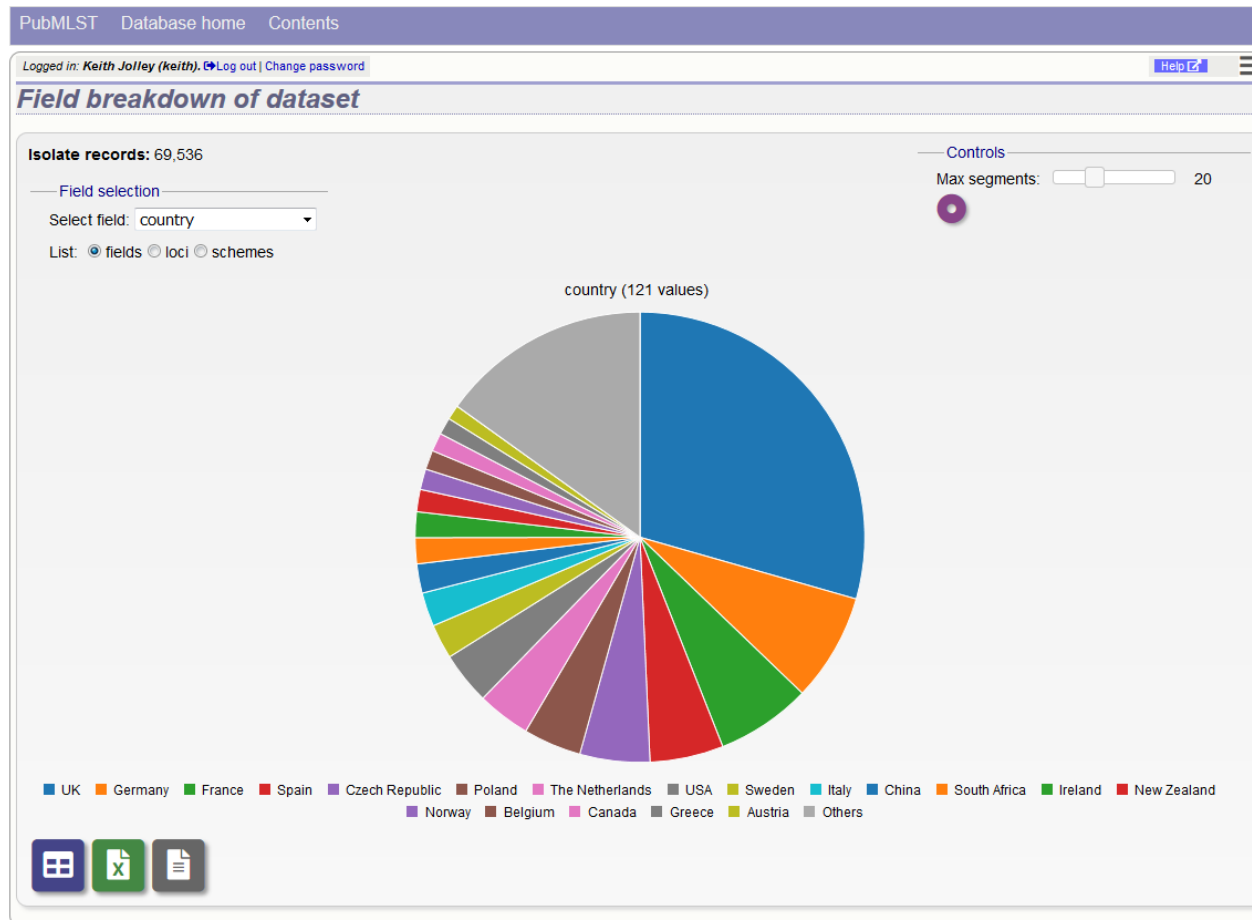
Export: Dataset | Contigs | Sequences

Third party: GrapeTree | iTOL | PhyloViz | Microreact

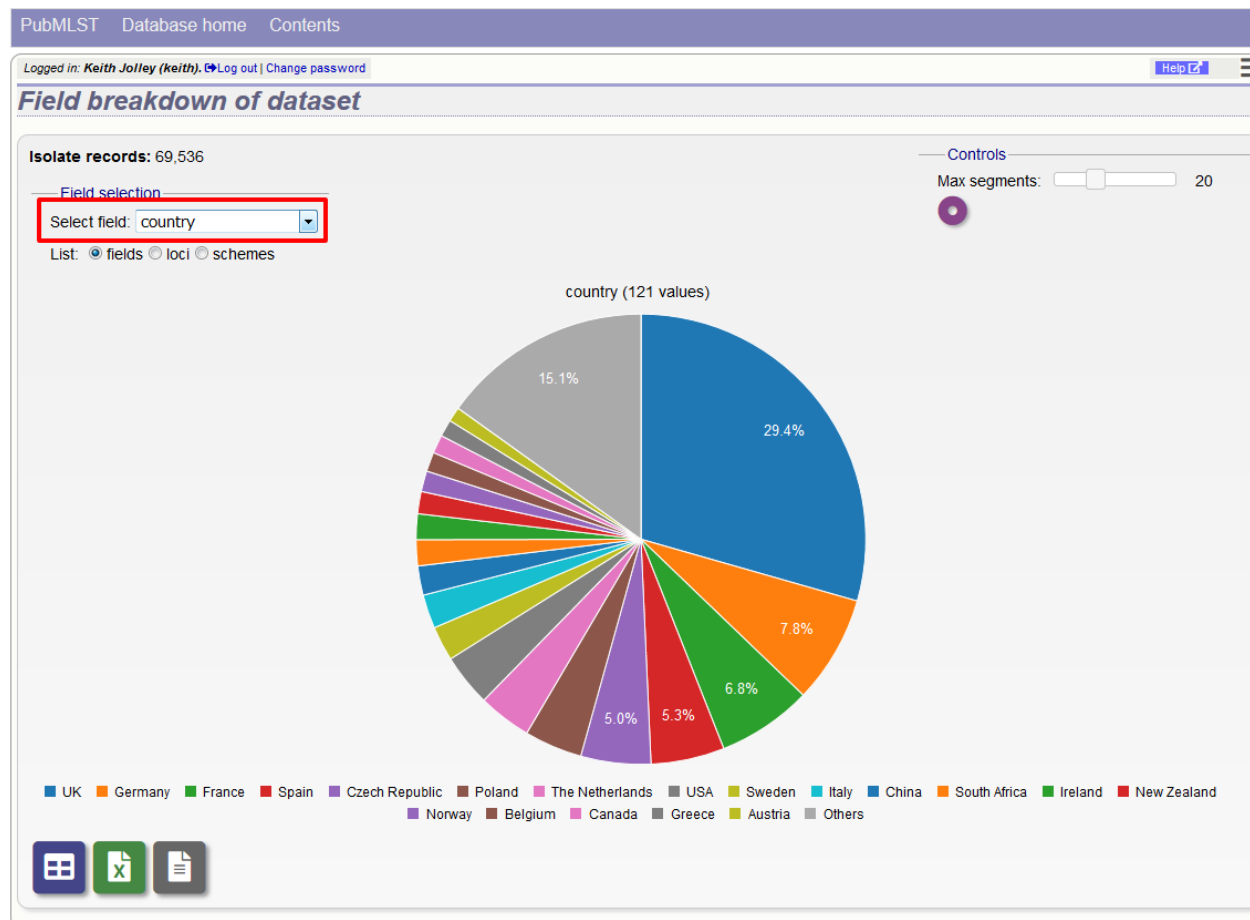
Page: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | > | Last

A chart will be displayed for the first field.

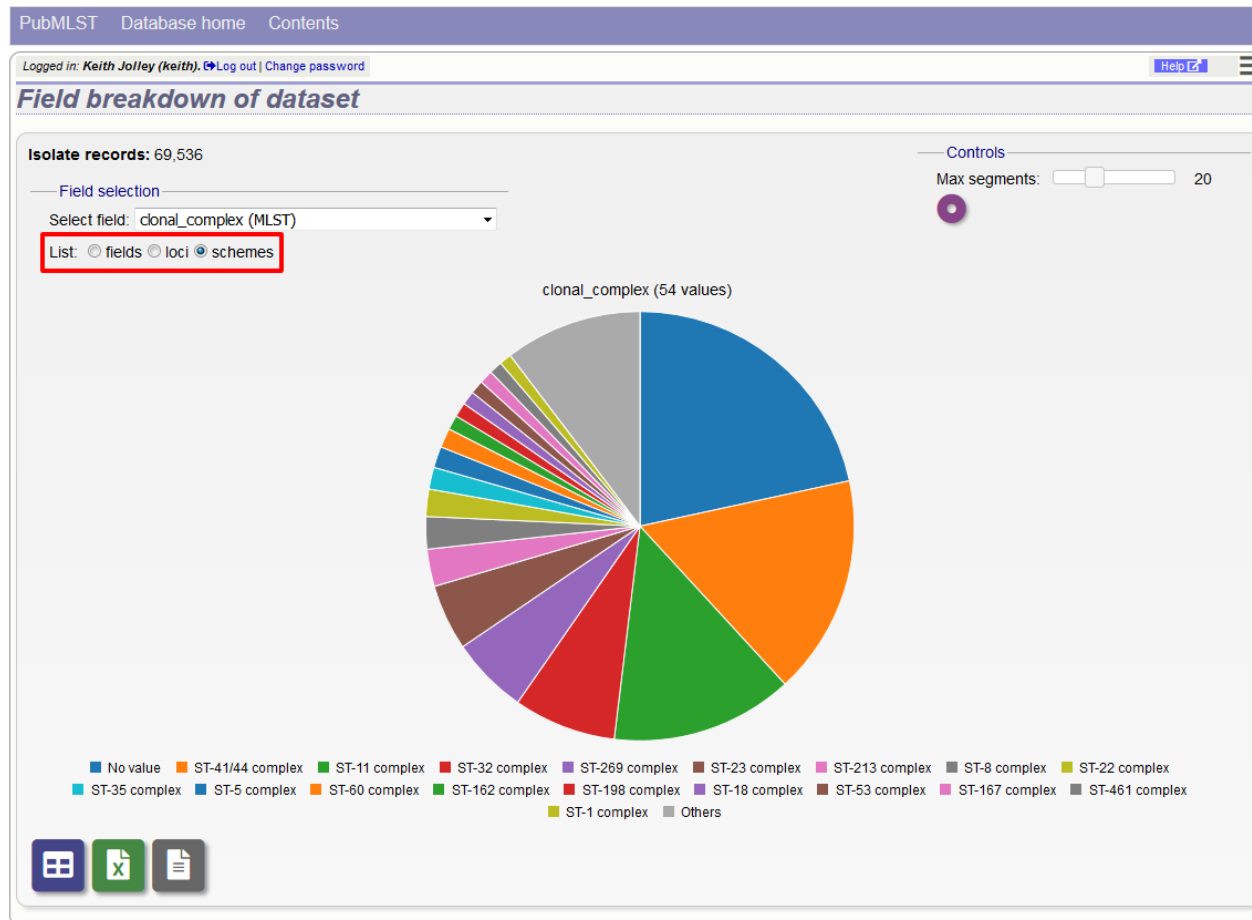




Other fields can be chosen by selecting them in the dropdown list box.



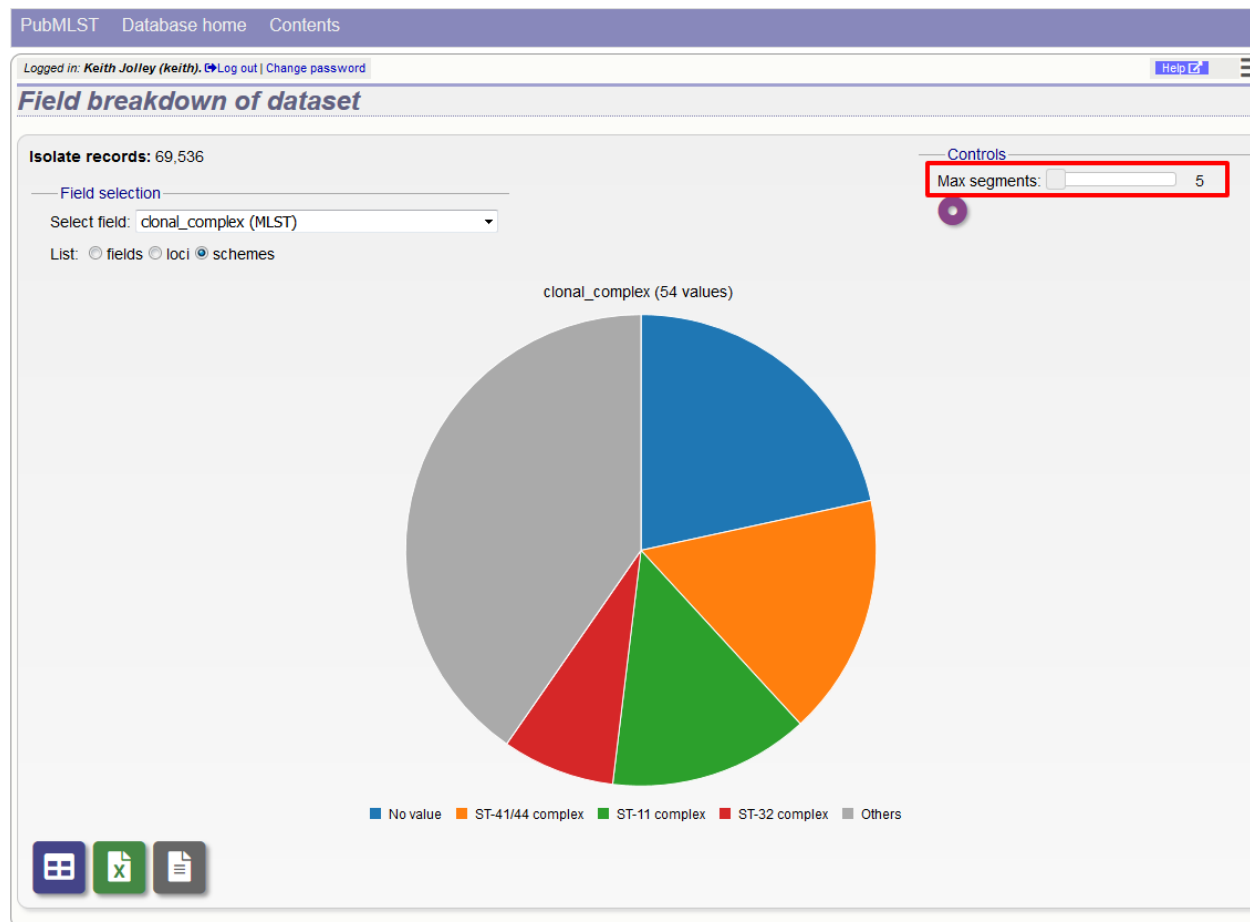
You can also breakdown loci and schemes by clicking the appropriate button. This will re-populate the dropdown list.



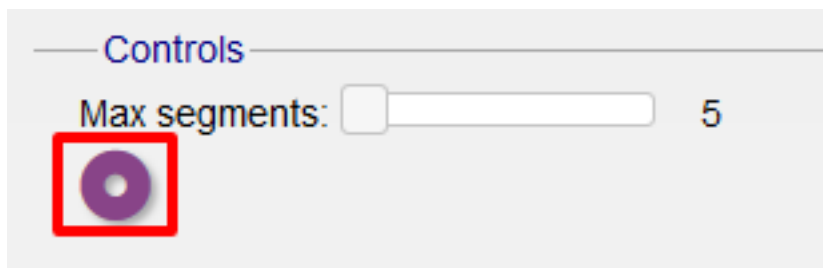
The charts are dynamic and you can manipulate some aspects of them using controls shown on the screen.

### 14.2.1 Pie charts

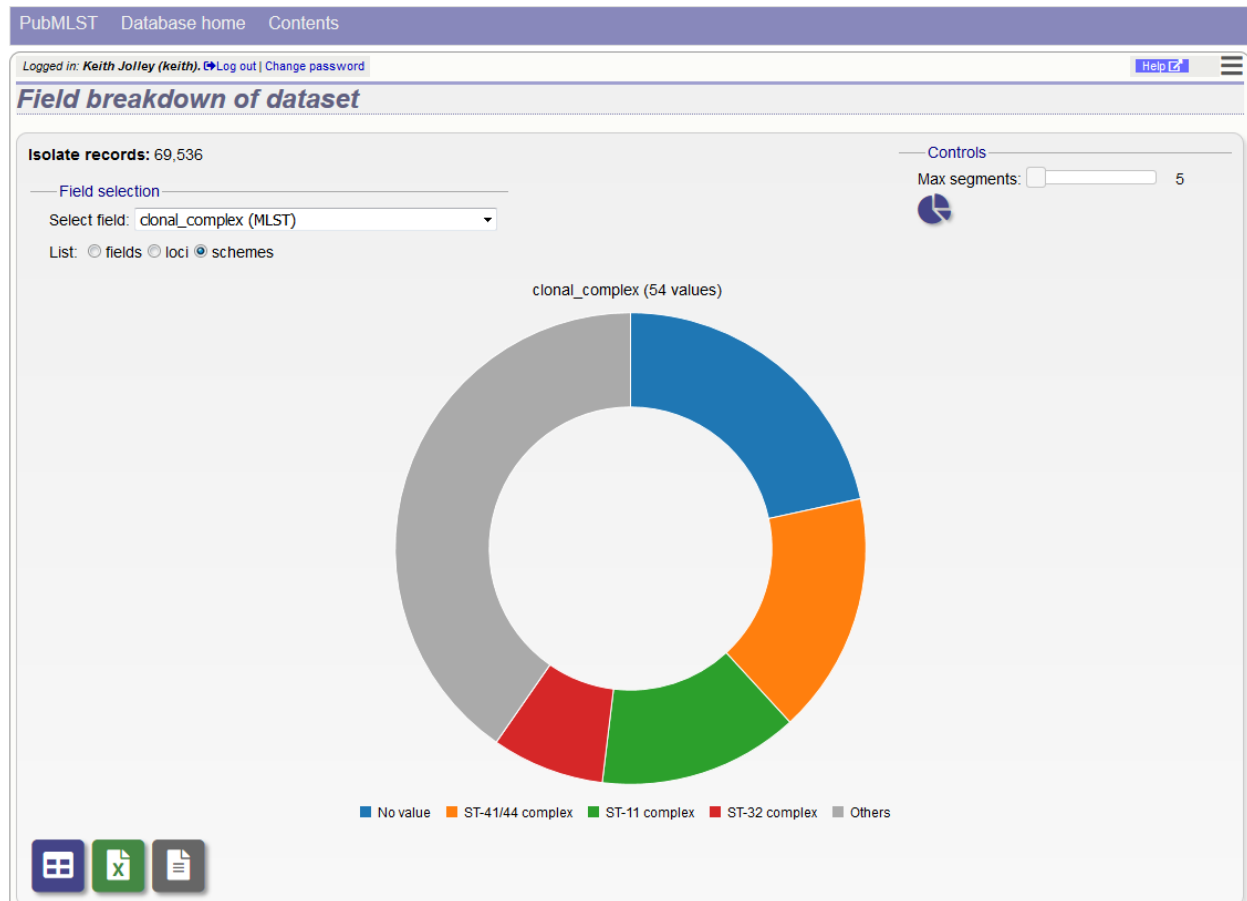
The maximum number of segments shown can be modified by sliding the 'Max segments' control. Low frequency values will be grouped in to a segment called 'Others'.



The chart can be transformed in to a donut chart by clicking the donut icon.



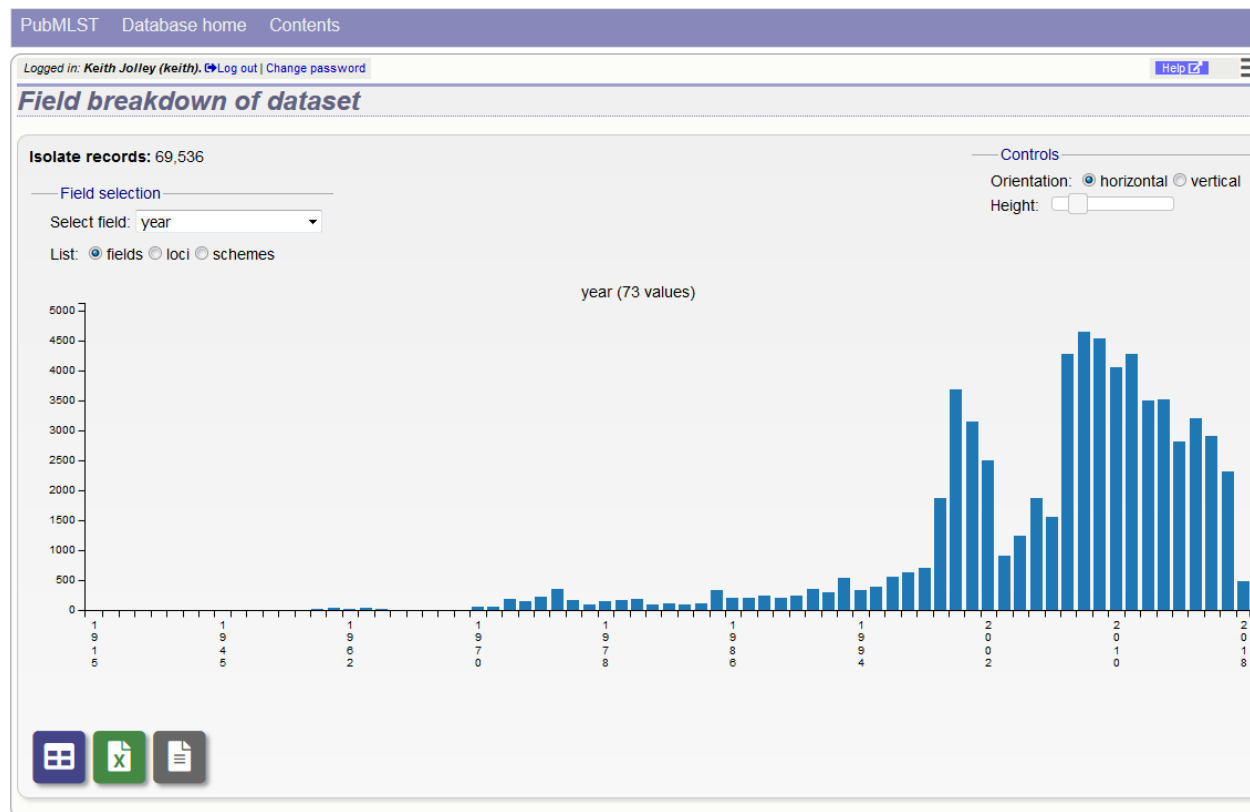
The icon changes to a pie chart image (clicking this will return to the pie chart).



Values can also be removed from the analysis by clicking their label in the legend below the chart. The percentages of the other values will be recalculated. Clicking the label again will re-add the value.

## 14.2.2 Bar charts

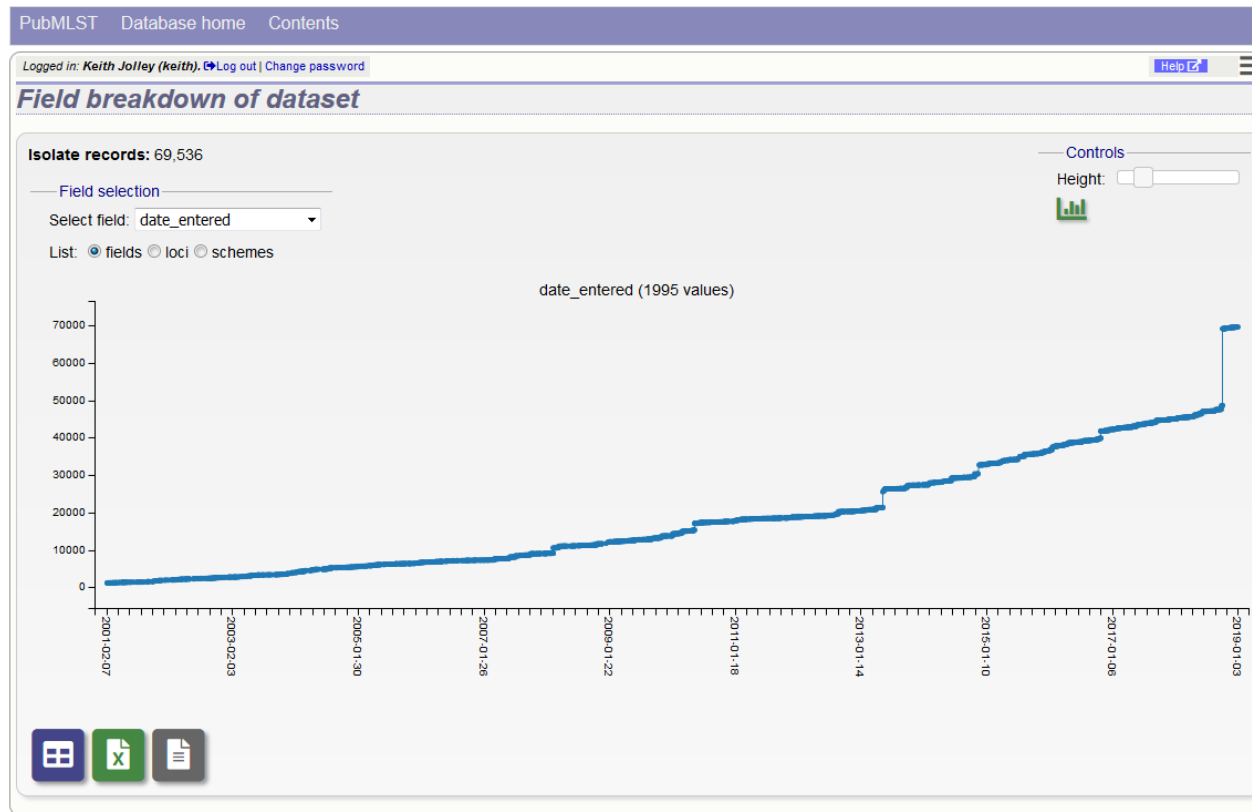
Integer fields will be displayed as a bar chart.



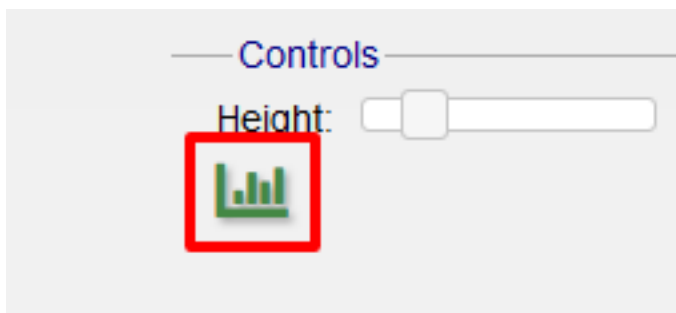
You can modify the height and the orientation of the chart using the controls.

### 14.2.3 Line charts

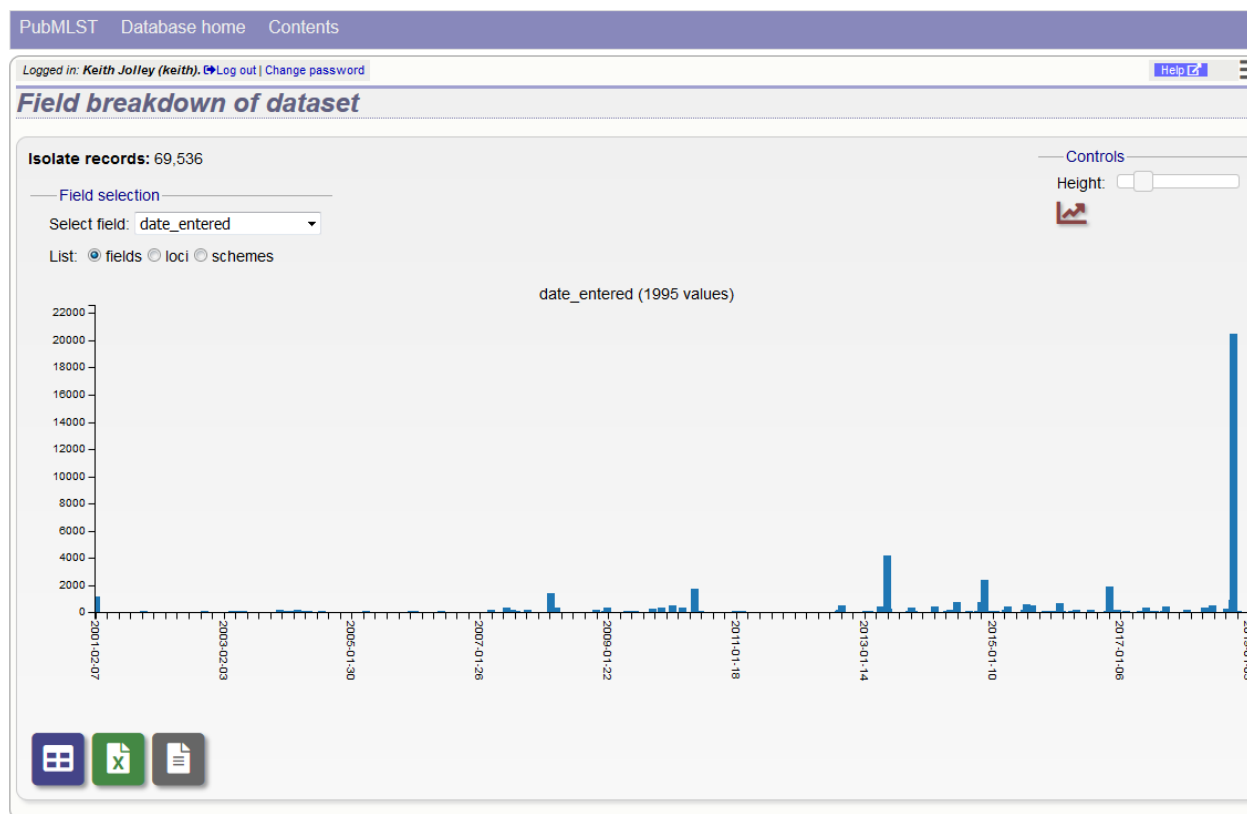
Date fields will be displayed as a line chart. By default this shows the cumulative values.



The chart can be converted in to a bar chart showing discrete values by clicking the bar chart icon.



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



#### 14.2.4 Summary tables

The field breakdown can be displayed as a summary table containing values and percentages of all values. This can be selected by clicking the table icon below the displayed chart.



The table can be re-ordered by clicking any of the headings.



PubMLST Database home Contents

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

### Breakdown by country

121 values.

country	Frequency	Percentage
UK	20460	29.42
Germany	5390	7.75
France	4744	6.82
Spain	3680	5.29
Czech Republic	3497	5.03
Poland	2897	4.17
The Netherlands	2665	3.83
USA	2618	3.76
Sweden	1730	2.49
Italy	1693	2.43
China	1457	2.10
South Africa	1300	1.87
Ireland	1283	1.85
New Zealand	1110	1.60
Norway	1033	1.49
Belgium	964	1.39
Canada	922	1.33
Greece	850	1.22
Austria	737	1.06
Brazil	651	0.94
Japan	604	0.87
Unknown	596	0.86
Finland	591	0.85
Portugal	590	0.85
Niger	580	0.83
Burkina Faso	570	0.82
Denmark	530	0.76
Russia	497	0.71
Cuba	432	0.62
Slovenia	388	0.56
Australia	321	0.46

The same table can be exported as an Excel file by clicking the Excel icon.

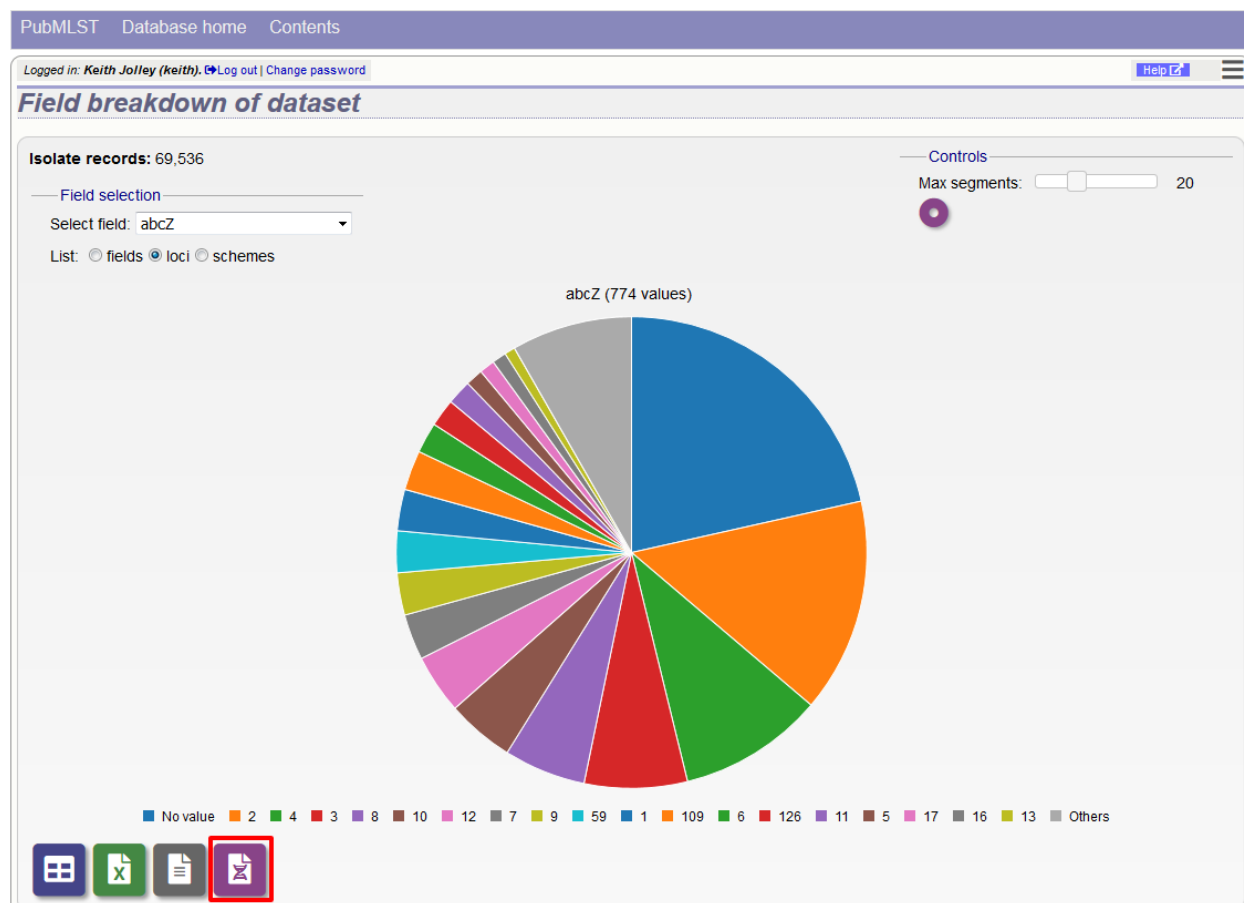


Alternatively, it can be exported as a tab-delimited text file by clicking the text file icon.



### 14.2.5 Exporting allele sequences

If a locus breakdown is being display, you can choose to export the allele sequences in FASTA format by clicking the FASTA file icon.



## 14.3 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.

6	MU028220		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		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	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 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

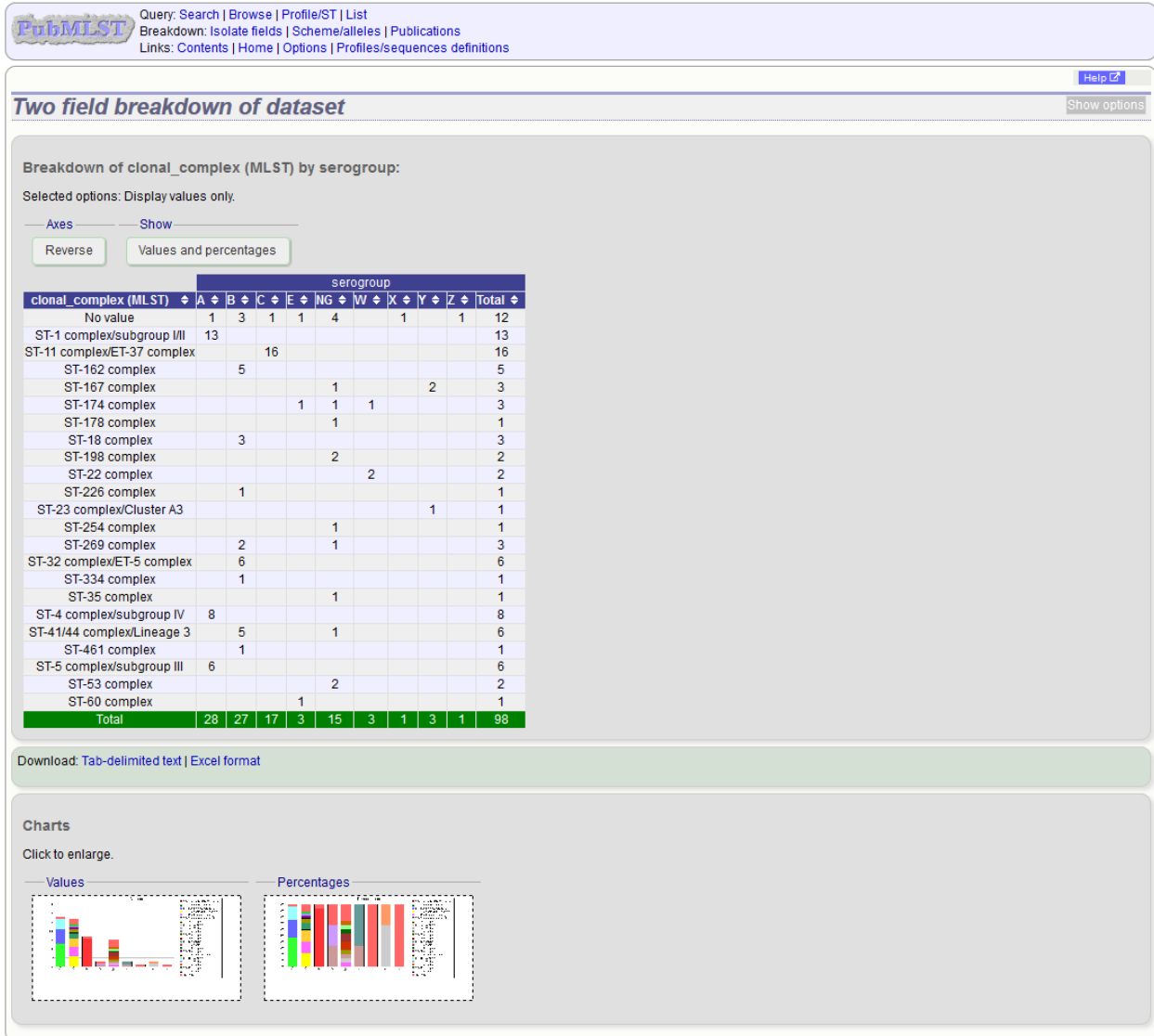
☒ dataset  
☐ row  
☐ column

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.

**Note:** The job will be submitted to the offline job queue if the query returns 10,000 or more isolates. In this case, the buttons to reverse the axes or to change whether values or percentages are shown will not be available.

## 14.4 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 - XMTA / 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	12UM	B43; Z1035	Pakistan	1997	meningitis and septicaemia	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	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: 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.

PubMLST 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  
 All None Paste list

**Loci**  
 16S\_rDNA  
 16S\_rRNA (SSU\_rRNA)  
 23S\_rRNA  
 abcZ  
 abcZ (NEIS1015)  
 aceF (NEIS1279)  
 ackA2 (NEIS1727)  
 acnA (NEIS1729)  
 All None Paste list

**Schemes**  
☒ All loci  
☒ Capsule  
☒ Genetic Information Proc  
☒ Genomic islands  
☒ Lineage Schemes  
☒ Metabolism  
☒ Pilin  
 ...

**Options**  
☐ Calculate %GC

Filter by  
 Sequence method:   
 Project:   
 Experiment:

Action  
 Reset Submit

If there are fewer than 100 isolates selected, the table will be generated immediately. Otherwise it will be submitted to the job queue.

A table of sequence bin stats will be generated.

PubMLST 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  
 All None Paste list

**Loci**  
 16S\_rDNA  
 16S\_rRNA (SSU\_rRNA)  
 23S\_rRNA  
 abcZ  
 abcZ (NEIS1015)  
 aceF (NEIS1279)  
 ackA2 (NEIS1727)  
 acnA (NEIS1729)  
 All None Paste list

**Schemes**  
☒ All loci  
☒ Capsule  
☒ Capsule Region A  
☒ Capsule Region A  
☒ Capsule Region A  
☒ Capsule Region A  
☒ Capsule Region A  
 ...

**Options**  
☐ Calculate %GC

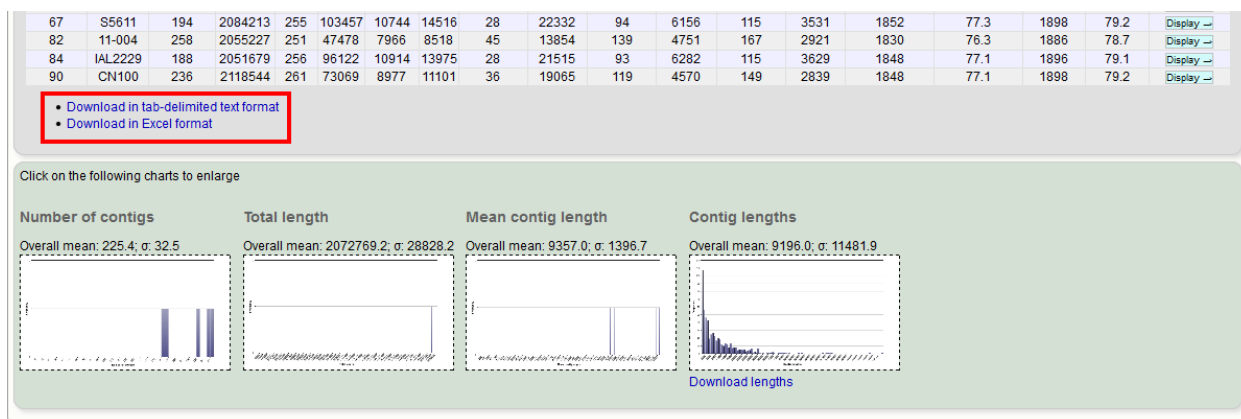
Filter by  
 Sequence method:   
 Project:   
 Experiment:

Action  
 Reset Submit

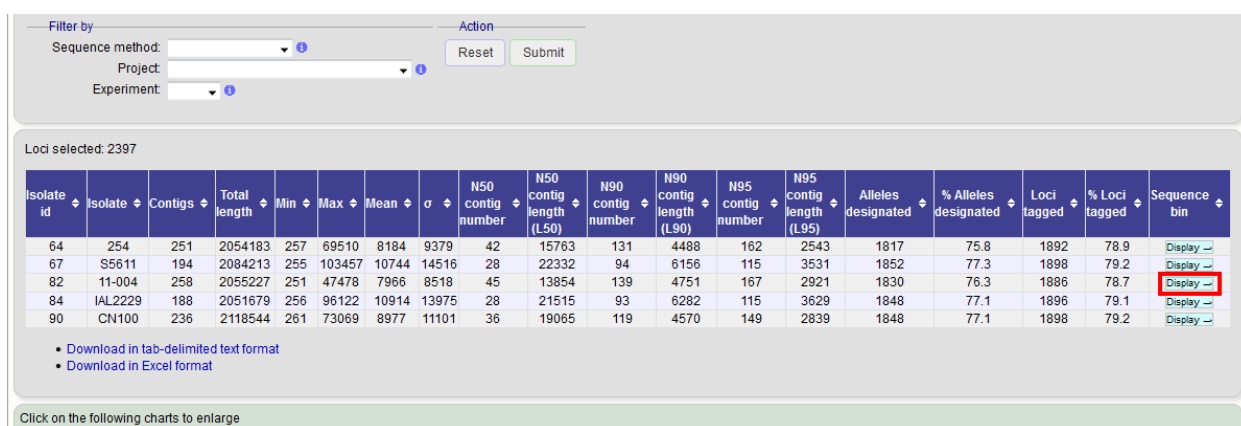
Loci selected: 2397

Isolate id	Isolate	Contigs	Total length	Min	Max	Mean	$\sigma$	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	<a href="#">Display</a>
67	S5611	194	2084213	255	103457	10744	14516	28	22332	94	6156	115	3531	1852	77.3	1898	79.2	<a href="#">Display</a>
82	11-004	258	2055227	251	47478	7966	8519	45	13854	139	4751	167	2921	1830	76.3	1886	78.7	<a href="#">Display</a>
84	IAL2229	188	2051679	256	96122	10914	13975	28	21515	93	6282	115	3629	1848	77.1	1896	79.1	<a href="#">Display</a>

You can choose to export the data in tab-delimited text or Excel formats by clicking the appropriate link at the bottom of the table.



*Sequence bin records* can also be accessed by clicking the ‘Display’ button for each row of the table.



## 14.5 Genome comparator

Genome Comparator is an optional plugin that can be enabled for specific databases. It is used to compare whole genome data of isolates within the database using either the database defined loci or the coding sequences of an annotated genome as the comparator.

Output is equivalent to a whole genome MLST profile, a distance matrix calculated based on allelic differences and a NeighborNet graph generated from this distance matrix.

Genome Comparator can be accessed on databases where it is enabled from the contents page by clicking the ‘Genome Comparator’ link.



Query: [Search](#) | [Browse](#) | [Profile/ST](#) | [List](#)  
 Breakdown: [Isolate fields](#) | [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 ‘Genome Comparator’ button at the bottom of the results table. Isolates with sequence data returned in the query will be automatically selected within the Genome Comparator interface.

19025	M10 240481	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	1831		22	9	F3-3
19026	M10 240482	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	3754	ST-41/44 complex/Lineage 3	7-2	4	F5-1
19027	M10 240484	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	1097	ST-41/44 complex/Lineage 3	17-1	23	F1-5
19028	M10 240485	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	275	ST-269 complex	22	9	F5-12
19029	M10 240487	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	1049	ST-269 complex	19-1	15-11	F5-1
19030	M10 240489	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	461	ST-461 complex	19-2	13-1	F3-9
19031	M10 240490	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	1161	ST-269 complex	22	9	F5-12
19032	M10 240498	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	41	ST-41/44 complex/Lineage 3	7-2	4	F1-5
19958	M10 240476	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	9812	ST-213 complex	22	14	F5-5
19959	M10 240499	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	275	ST-269 complex	22	9	F5-12
19960	M10 240500	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	41	ST-41/44 complex/Lineage 3	7-2	4	F1-5
19961	M10 240502	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	340	ST-41/44 complex/Lineage 3	7-2	4	F1-5
19962	M10 240503	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	Y	23	ST-23 complex/Cluster A3	5-1	2-2	F1-96
19963	M10 240505	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	Y	1655	ST-23 complex/Cluster A3	5-1	10-10	F4-1
19964	M10 240507	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	Y	183	ST-23 complex/Cluster A3	21	16-5	deleted
19965	M10 240508	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	1575		7-2	13-1	F1-7
19966	M10 240511	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	4713		22	9	F5-12
19967	M10 240512	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	41	ST-41/44 complex/Lineage 3	7-1	1	F1-5
19968	M10 240514	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	W	11	ST-11 complex/ET-37 complex	5	2	F1-1
19969	M10 240515	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B	269	ST-269 complex	5-1	2-2	F5-1
19970	M10 240520	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	Y	1655	ST-23 complex/Cluster A3	5-1	10-1	F4-1

Analysis tools:

Breakdown:

Analysis:

Export:

Page:

### 14.5.1 Analysis using defined loci

Select the isolate genomes that you wish to analyse. These will either be in a dropdown list or, if there are too many in the database, a text input where a list can be entered. You can also upload your own genomes for analysis - these should be either a single file in FASTA format (if you have just one genome), or a zip file containing multiple FASTA files. Select either the loci from the list or a set of schemes. Press submit.

PubMLST Database home Contents

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

### Genome Comparator - Neisseria PubMLST

Please select the required isolate ids and loci for comparison - use CTRL or SHIFT to make multiple selections in list boxes. In addition to selecting individual loci, you can choose to include all loci defined in schemes by selecting the appropriate scheme description. Alternatively, you can enter the accession number for an annotated reference genome and compare using the loci defined in that.

Isolates	User genomes	Loci	Include in identifiers	Schemes
<input type="checkbox"/> 662 <input type="checkbox"/> 663 <input type="checkbox"/> 664 <input type="checkbox"/> 665	Optionally include data not in the database. Upload FASTA file (or zip file containing multiple FASTA files - one per genome): <input type="button" value="Browse..."/> No file selected.	<input type="checkbox"/> 16S_rDNA <input type="checkbox"/> 16S_rRNA (SSU_rRNA) <input type="checkbox"/> 23S_rRNA <input type="checkbox"/> abcZ (NEIS1015) <input type="checkbox"/> aceF (NEIS1279) <input type="checkbox"/> ackA2 (NEIS1727) <input type="checkbox"/> acnA (NEIS1729)	<input type="checkbox"/> isolate <input type="checkbox"/> country <input type="checkbox"/> region <input type="checkbox"/> year <input type="checkbox"/> epidemiological year <input type="checkbox"/> age yr <input type="checkbox"/> age mth <input type="checkbox"/> sex <input type="checkbox"/> disease <input type="checkbox"/> source	<input type="checkbox"/> N. gonorrhoeae AMR <input type="checkbox"/> Pilin <input type="checkbox"/> Plasmids <input type="checkbox"/> Typing <input checked="" type="checkbox"/> MLST <input type="checkbox"/> Finotyping antigen

Reference genome	Parameters / options	Distance matrix calculation	Alignments	Core genome analysis
Enter accession number: <input type="text"/> or choose annotated genome: <input type="text"/> or upload Genbank/EMBL/FASTA file: <input type="button" value="Browse..."/> No file selected.	Min % identity: <input type="text" value="70"/> Min % alignment: <input type="text" value="50"/> BLASTN word size: <input type="text" value="20"/>	With incomplete loci: <input type="radio"/> Completely exclude from analysis <input type="radio"/> Treat as distinct allele <input checked="" type="radio"/> Ignore in pairwise comparison <input checked="" type="checkbox"/> Exclude paralogous loci	<input type="checkbox"/> Produce alignments <input checked="" type="checkbox"/> Include ref sequences in alignment <input type="checkbox"/> Align all loci (not only variable) Aligner: <input type="text" value="MAFFT"/>	Core threshold (%): <input type="text" value="90"/> <input type="checkbox"/> Calculate mean distances

Filter by: Sequence method:   
 Project:   
 Experiment:

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.

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

Toggle: [i] Show options

This analysis has been submitted to the job queue.  
 Please be aware that this job may take a long time depending on the number of comparisons and how busy the server is.

[Follow the progress of this job and view the output.](#)

There will be a series of tables displaying variable loci, colour-coded to indicate allelic differences. Finally, there will be links to a distance matrix which can be loaded in to SplitsTree for further analysis and to a NeighborNet chart showing relatedness of isolates. Due to processing constraints on the web server, this NeighborNet is only calculated if 200 or fewer genomes are selected for analysis, but this can be generated in the stand-alone version of SplitsTree using the distance matrix if required.



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

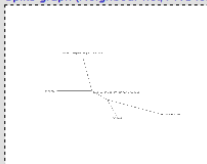
Locus	644 (L93/4286)	662 (2837)	663 (2839)	664 (2838)	665 (2845)	666 (2843)	667 (2842)	669 (2846)	670 (2840)	671 (2844)	672 (2847)	698 (FAM18)
adk	3	3	3	3	3	3	3	3	3	3	3	3
gdh	8	8	8	8	8	8	8	8	8	8	8	8
pgm	6	6	6	6	6	6	6	6	6	6	6	6

#### Unique strains

Unique strains: 5

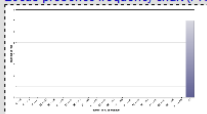
Strain 1	Strain 2	Strain 3	Strain 4	Strain 5
644 (L93/4286)	662 (2837)	666 (2843)	665 (2845)	669 (2846)
663 (2839)	670 (2840)	667 (2842)		
664 (2838)	671 (2844)			
698 (FAM18)	672 (2847)			

- 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.

## 14.5.2 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.

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.

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

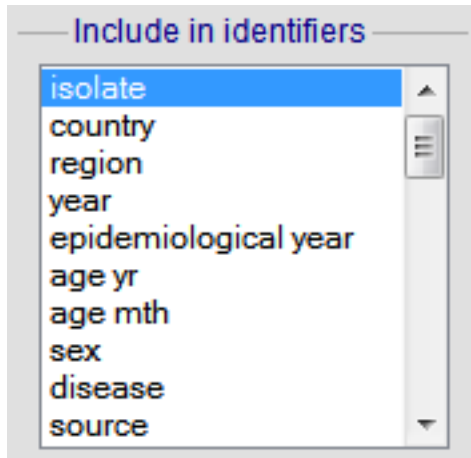
**All loci**

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
fbp   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	methylionyl-RNA 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

### 14.5.3 Include in identifiers fieldset

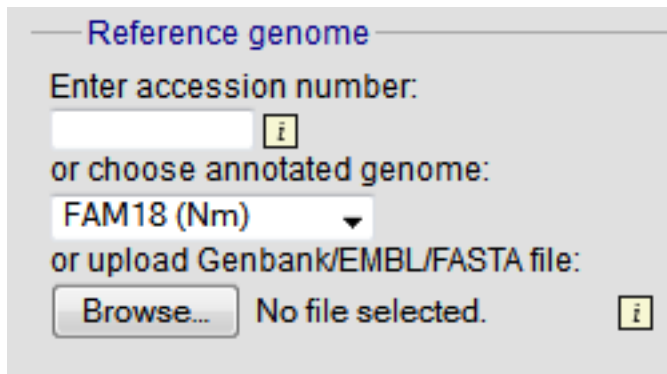
This selection box allows you to choose which isolate provenance fields will be included in the results table. This does not affect the output of the alignments as taxa names are limited in length by the alignment programs.



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

### 14.5.4 Reference genome fieldset

This section allows you to choose a reference genome to use as the source of comparator sequences.

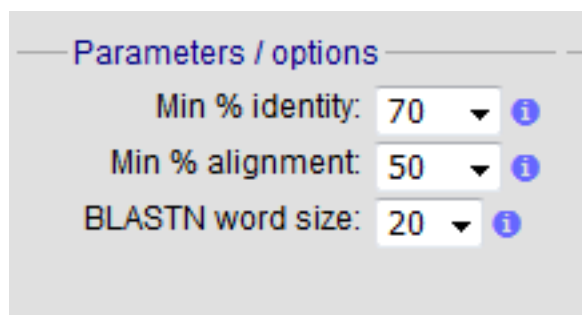


There are three possibilities here:



1. Enter accession number - Enter a Genbank accession number of an annotated reference and Genome Comparator will automatically retrieve this from Genbank.
2. Select from list - The administrator may have selected some genomes to offer for comparison. If these are present, simply select from the list.
3. Upload genome - Click 'Browse' and upload your own reference. This can either be in Genbank, EMBL or FASTA format. Ensure that the filename ends in the appropriate file extension (.gb, .embl, .fas) so that it is recognized.



### 14.5.5 Parameters/options fieldset



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



Parameters / options

Min % identity: 70  

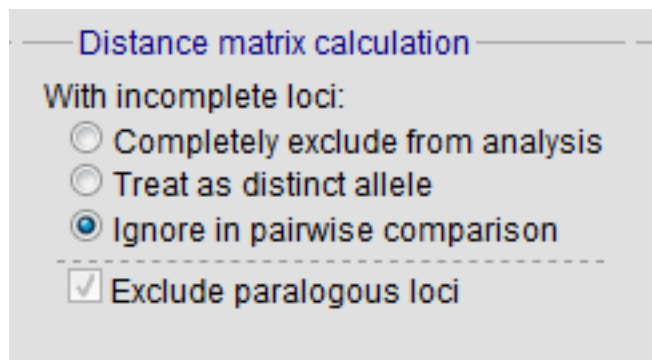
Min % alignment: 50  

BLASTN word size: 20  

- Min % identity - This sets the threshold identity that a matching sequence has to be in order to be considered (default: 70%). Only the best match is used.
- Min % alignment - This sets the percentage of the length of reference allele sequence that the alignment has to cover in order to be considered (default: 50%).
- BLASTN word size - This is the length of the initial identical match that BLAST requires before extending a match (default: 20). Increasing this value improves speed at the expense of sensitivity. The default value gives good results in most cases. The default setting used to be 15 but the new default of 20 is almost as good (there was 1 difference among 2000 loci in a test run) but the analysis runs twice as fast.

### 14.5.6 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

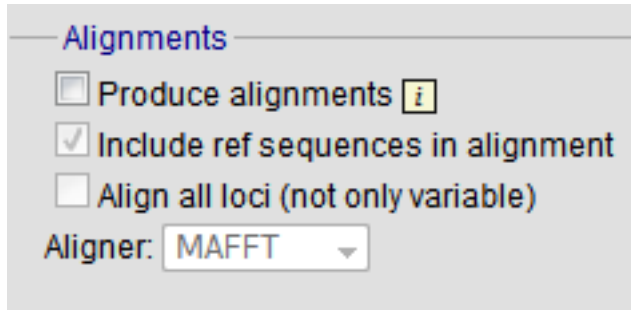
For incomplete loci, i.e. those that continue beyond the end of a contig so are incomplete you can:

- Completely exclude from analysis - Any locus that is incomplete in at least one isolate will be removed from the analysis completely. Using this option means that if there is one bad genome with a lot of incomplete sequences in your analysis, a large proportion of the loci may not be used to calculate distances.
- Treat as a distinct allele - This treats all incomplete sequences as a specific allele 'I'. This varies from any other allele, but all incomplete sequences will be treated as though they were identical.
- Ignore in pairwise comparison (default) - This is probably the best option. In this case, incomplete alleles are only excluded from the analysis when comparing the particular isolate that has it. Other isolates with different alleles will be properly included. The effect of this option will be to shorten the distances of isolates with poorly sequenced genomes with the others.


Paralogous loci, i.e. those with multiple good matches, can be excluded from the analysis (default). This is the safest option since there is no guarantee that differences seen between isolates at paralogous loci are real if the alternative matches are equally good. NB: Loci are also only classed as paralogous when the alternative matches identify different sequences, otherwise multiple contigs of the same sequence region would result in false positives.

### 14.5.7 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)

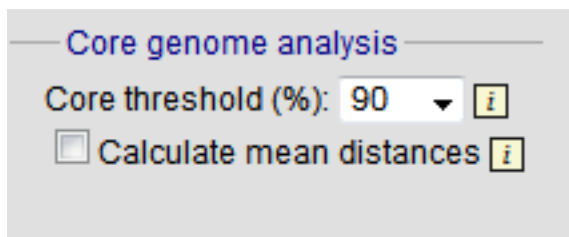
Aligner: MAFFT ▼

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 longer to run.
- Include ref sequences in alignment - When doing analysis using an annotated reference, selecting this will include the reference sequence in the alignment files.
- Align all loci - By default, only loci that vary among the isolates are aligned. You may however wish to align all if you would like the resultant XMFA and FASTA files to include all coding sequences.
- Aligner - There are currently two choices of alignment algorithm (provided they have both been installed)
  - MAFFT (default) - This is the preferred option as it is significantly quicker than MUSCLE, uses less memory, and produces comparable results.
  - MUSCLE - This was originally the only choice. It is still included to enable previous analyses to be re-run and compared but it is recommended that MAFFT is used otherwise.


### 14.5.8 Core genome analysis fieldset

This section enables you to modify the inclusion threshold used to calculate whether or not a locus is part of the core genome (of the dataset).



**Core genome analysis**

Core threshold (%): 90 ▼ 

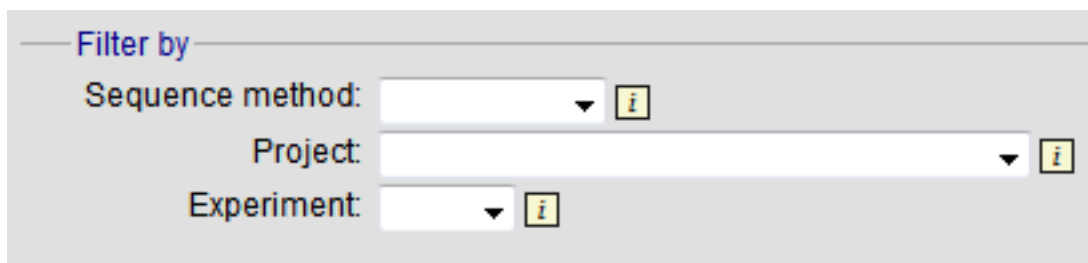
☐ Calculate mean distances 

The default setting of 90% means that a locus is counted as core if it appears within 90% or more of the genomes in the dataset.


There is also an option to calculate the mean distance among sequences of the loci. Selecting this will also select the option to produce alignments.


### 14.5.9 Filter fieldset


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



Filter 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.

## 14.5.10 Understanding the output

### 14.5.10.1 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.

### 14.5.10.2 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.

## 14.6 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](#) | [Next](#) | [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.

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

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:  Paste sequence:

Include in results table:

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.

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

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:  Paste sequence:

Include in results table:

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	<a href="#">extract</a>	→	0.0	807
182004	36775	37239	<a href="#">extract</a>	→	0.0	807
182318	19090	19554	<a href="#">extract</a>	→	0.0	807
215673	4534	4998	<a href="#">extract</a>	←	0.0	807
8	1363524	1363988	<a href="#">extract</a>	→	0.0	785
182380	5559	6023	<a href="#">extract</a>	→	0.0	807
182815	19783	20247	<a href="#">extract</a>	→	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	Solexa
start	19090
end	19554
length	465
orientation	→
translation	

```

CGGACGGGT TGAACAGC ATCCGGAATA GCGCGAATA GCGCGGCCA AACTGCGCGA ATTCGCGG TCGCGGCTTG GCGCGGCG GAACAAATTT GAAGCCTTGG GCGGACGCGA TGCCGCGGTT GCGCGCTCGG
GCGCATTTGAA AACGCTGGCG GCAAGCCTGA ATAAATCGC CAACGACATC CGCTGGCTGG CAAGCGGCCC GCGCTGCGGT TTGGGCGAAA TCAGAAATCCC CAGAAACGAG CGGGTTTGGT CCATCATGCG GGGCAAGTC
AACCGACCC AATCGGAAGC GATGACCATG GTGTGCTGCC AAGTGTTCGG CAACGACGTT ACCATCGGTA TGGCGGCGCG GTGCGGCAAT TTCGAGCTGA ACGTCTATAT GCCGTCATC GCGTACAACC TCTTGCATC
CATCGGCTG TTGGGCGAGC CGTGCAACAG CTTCAACGAA CACTGCGCGG TCGGCAITGA ACCGCTACCG GAAAAAATCG ACTATTTCCT GCACCATTCG CTGATGCTCG TTACCGCGTT AAACCGCAAA ATCGGTTACG
AAAACCGCC CAAAGTCGCC AAAACCGCCT ACAGAAACAA CAATCGTGG CCGGAACCG CCGTTGAGTT GGGCTTCTG ACGGGCGAAG AATTTCAGCA ACTGG

```

Translation:

```

R H G F E Q P S R I R R K S R R Q T R R I V R L A V C Q R A E Q I * 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 CGGACGGGT TGAACAGC ATCCGGAATA GCGCGAATA GCGCGGCCA AACTGCGCGA ATTCGCGG TCGCGGCTTG GCGCGGCG GAACAAATTT GAAGCCTTGG GCGGACGCGA TGCCGCGGTT GCGCGCTCGG
-----
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 GAAGCCTTGG GCGGACGCGA TGCCGCGGTT GCGCGCTT GCGGCGCG CACTGCGGCAAGCTTGAATTAATTCGCGGCAAGCATCCGCTGGCTGG
-----
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 CAAGCGGCGGCGCTGGGCTTGGGCGAATCAAAATCCCGAAGACGCGCGGTTTCGTCATCATGCGGCGAAGTCACCGCGCCCAATCGGAAGC
-----
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 A I S S * T S I C P S S F3
301 GATGACCATGCTGCTGCCAAGTGTTCGCGACGACGTTACCATCGGTATGGCGGCGCGCTCGGCGCAATTCGAGCTGAACGCTATATGCGCGTCATC
-----
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 GCGTACACCTCTTTCGAATCCATCCGCTGTGGGCGACGCGTGCAACAGCTTCAACGACACTGCGCGCTGCGCATTGAACCGCTACCGGAAATTCG
-----

```

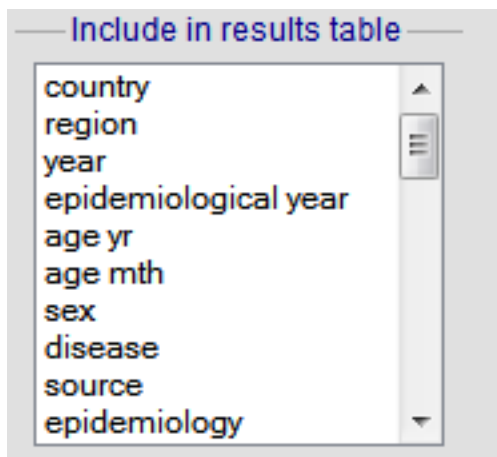
At the bottom of the results table are links to export the matching sequences in FASTA format, (optionall) including flanking sequences. You can also export the table in tab-delimited text or Excel formats.

87	35889	36353	465	7	0	184143	4004	5200	<a href="#">extract</a>	←	0.0	807
82	11-004	98.49	465	7	0	184297	19518	19982	<a href="#">extract</a>	→	0.0	807
84	IAL2229	98.49	465	7	0	184635	4530	4994	<a href="#">extract</a>	←	0.0	807
90	CN100	98.49	465	7	0	184814	5350	5814	<a href="#">extract</a>	←	0.0	807

[Download FASTA](#) | [FASTA with flanking](#) | [Table \(tab-delimited text\)](#) | [Excel format](#)

### 14.6.1 Include in results table fieldset

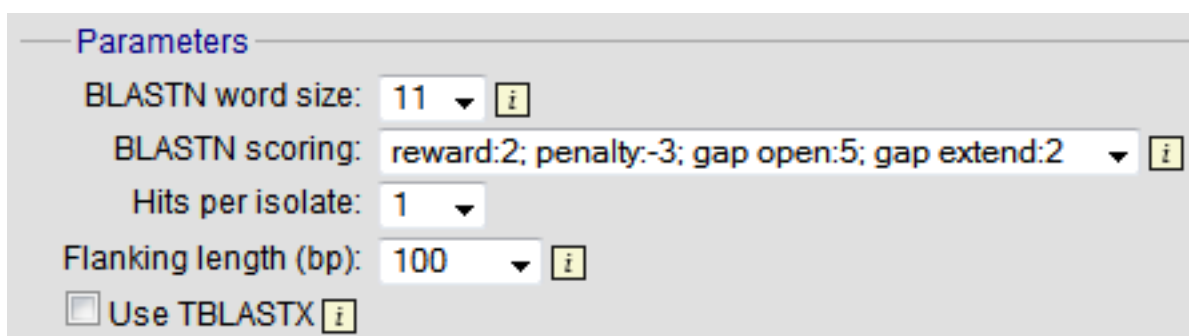
This selection box allows you to choose which isolate provenance fields will be included in the results table.



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

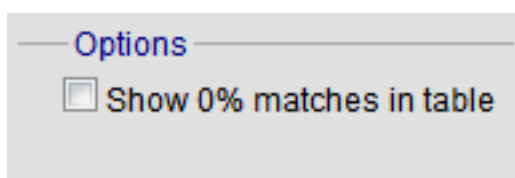
### 14.6.2 Parameters fieldset

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.

### 14.6.3 No matches





Click this option to create a row in the table indicating that a match was not found. This can be useful when screening a large number of isolates.


### 14.6.4 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.

## 14.7 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)

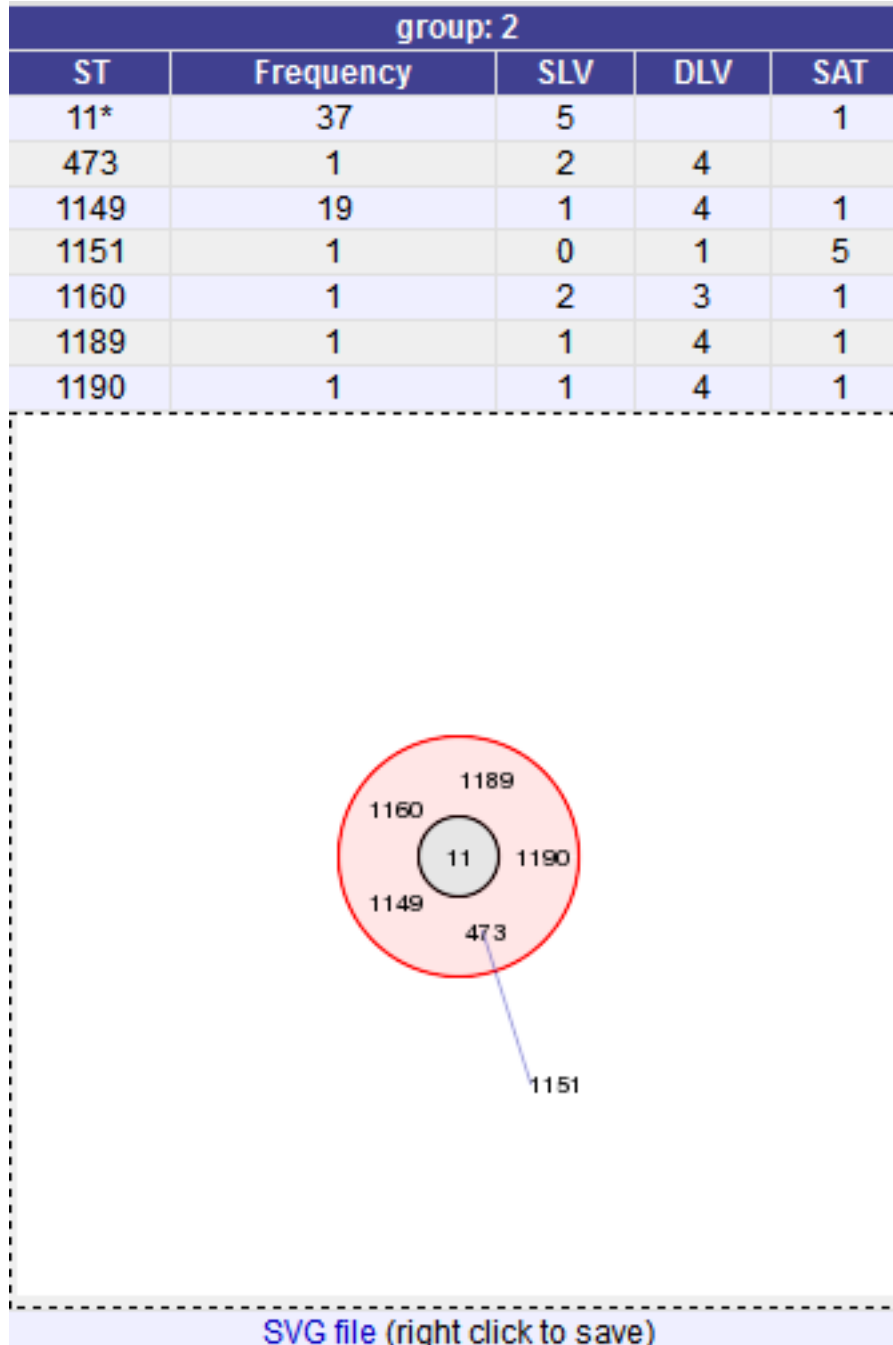
Action

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.

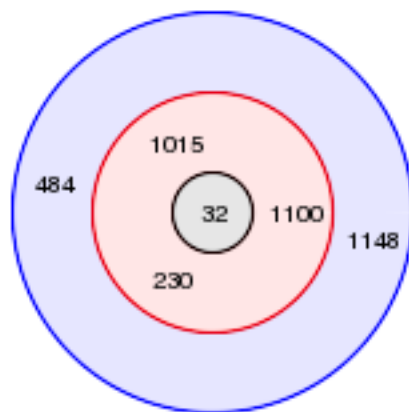
Click Submit.

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.

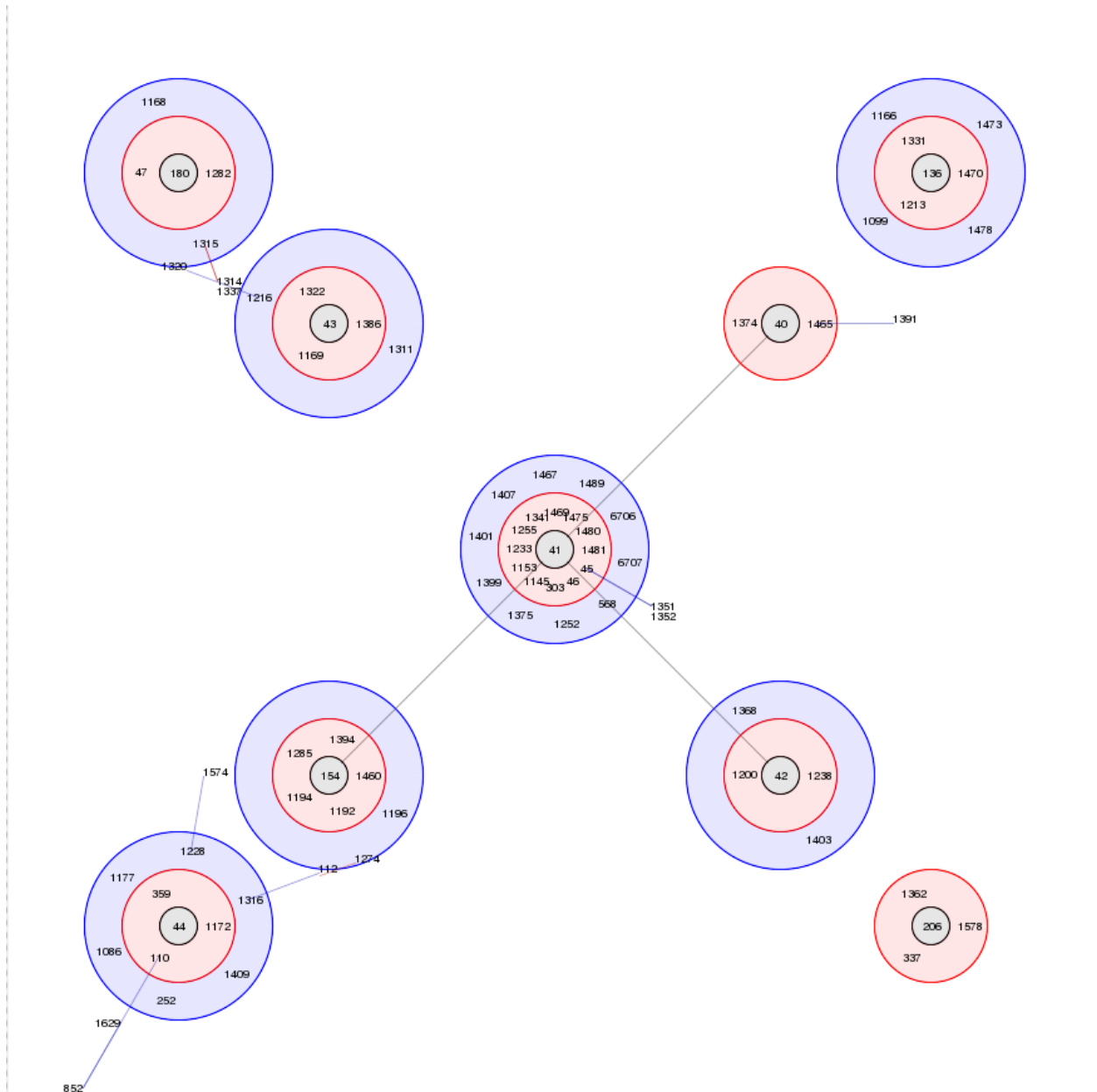
group: 6				
ST	Frequency	SLV	DLV	SAT
32*	2	3	2	
230	1	1	3	1
484	1	0	3	2
1015	1	1	4	
1100	1	1	2	2
1148	1	0	4	1



[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.





## 14.8 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)	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											
10	6748	B73; Z1073	Czech Republic	1984	invasive (unspecified/other)	Neisseria meningitidis	W	114	ST-22 complex			
11	129	B92; Z1092	Germany	1964	invasive (unspecified/other)	Neisseria meningitidis	A	1	ST-1 complex/subgroup I/II	18-1	3	F5-1
12	0090/89		Czech Republic	1989	invasive (unspecified/other)	Neisseria meningitidis	B	1015	ST-32 complex/ET-5 complex	7	16	F3-6
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](#)

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

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_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.

## 14.9 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.

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

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

☐ isolate ☐ MLEE designation ☐ rifampicin range

☐ aliases ☐ serotype ☐ ciprofloxacin

☐ country ☐ sero subtype ☐ ciprofloxacin range

☐ continent ☐ ET no ☐ pending assembly

☐ region ☐ penicillin ☐ assembly status

☐ year ☐ penicillin range ☐ ENA accession

☐ epidemiological year ☐ amoxicillin ☐ private project

☐ age yr ☐ sulphonamide ☐ comments

☐ age mth ☐ ceftriaxone ☐ sender

☐ sex ☐ ceftriaxone range ☐ curator

☐ disease ☐ chloramphenicol ☐ date entered

☐ source ☐ chloramphenicol range ☐ datestamp

☐ epidemiology ☐ cefotaxime

☐ species ☐ cefotaxime range

**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: [i]

### 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](#)

## 14.10 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-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 5 6 7 8 > Last

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

PubMLST 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: [i]

☒ 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.


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


[Help](#) 
[Toggle](#) 

## Polymorphic site analysis

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.

17 alleles included in analysis. 81 polymorphic sites found.

Key: 0 - 10 | >10 - 20 | >20 - 30 | >30 - 40 | >40 - 50 | >50 - 60 | >60 - 70 | >70 - 80 | >80 - 90 | >90 - 100



Nucleotide frequencies

Position	Nucleotide									
	A	C	G	T	-	%A	%C	%G	%T	%-
9	0	14	0	3	0		82.35		17.65	
24	1	0	16	0	0	5.88		94.12		
27	0	10	0	7	0		58.82		41.18	
28	4	0	13	0	0	23.53		76.47		
30	16	0	1	0	0	94.12		5.88		
36	0	14	0	3	0		82.35		17.65	

## 14.11 Gene Presence

The Gene Presence analysis tool will determine whether loci are present or absent, incomplete, have alleles designated, or sequence regions tagged for selected isolates and loci. If a genome is present and a locus designation not set in the database, then the presence and completion status are determined by scanning the genomes. The results can be displayed as interactive pivot tables or a heatmap. The analysis is limited to 500,000 data points (locus x isolates).

The Gene Presence tool can be accessed from the contents page by clicking the 'Gene Presence' link.



PubMLST Database home Contents

Log in

## 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 or browse database
- Search by combinations of loci (profiles)

**Projects**

- Main public projects
- Your projects

**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: 48,296
- Last updated: 2019-01-28
- Defined field values
- Update history
- About BIGSdb

**Breakdown**

- Single field
- Two field
- Unique combinations
- Publications
- Sequence bin

**Export**

- Export dataset
- Contigs
- Sequences - XMTA / concatenated FASTA formats

**Miscellaneous**

- Description of database fields

**Analysis**

- Codon usage
- Gene presence**
- Genome comparator
- BLAST
- Species identification

**Third party tools**

- GrapeTree - Visualization of genomic relationships
- ITOL - Phylogenetic trees with data overlays
- PhyloViz - Visualization and phylogenetic inference
- Microreact - Open data visualization and sharing for genomic epidemiology

Alternatively, it can be accessed following a query by clicking the ‘Gene Presence’ button at the bottom of the results table. Isolates returned from the query will be automatically selected within the plugin interface.

ST	Accession	Country	Year	Isolate type	Neisseria meningitidis	ST	Complex	Other
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	Ghana	1973	invasive (unspecified/other)	Neisseria meningitidis	A	4	ST-4 complex 7 13-1 F1-5
20	5	Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex
21	6	Germany	1999	carrier	Neisseria meningitidis	NG	198	ST-198 complex
22	7	Germany	1999	carrier	Neisseria meningitidis	E	60	ST-60 complex
23	8	Germany	1999	carrier	Neisseria meningitidis	B	32	ST-32 complex
24	S4355	Denmark	1974	invasive (unspecified/other)	Neisseria meningitidis	A	5	ST-5 complex 5-1 9 F3-1
25	9	Germany	1999	carrier	Neisseria meningitidis	B	930	ST-334 complex

Analysis tools:

Breakdown: Fields Two Field Combinations Polymorphic sites Publications Sequence bin

Analysis: BURST Codons **Gene Presence** Genome Comparator BLAST rMLST species id

Export: Dataset Contigs Sequences

Third party: GrapeTree ITOL PhyloViz Microreact

Page: 1 2 3 4 > Last

Select the isolates to include. Analysis can be performed on any selection of loci, or more conveniently, you can select a scheme in the scheme selector to include all loci belonging to that scheme.

The parameters of the BLAST query used to determine presence or absence can be modified, but in most cases the

default options should work well. Click ‘Submit’ to start the analysis.

PubMLST Database home Contents

Log in Toggle: 1

### Gene Presence - *Neisseria* PubMLST

Please select the required isolate ids and loci for comparison - use CTRL or SHIFT to make multiple selections in list boxes. In addition to selecting individual loci, you can choose to include all loci defined in schemes by selecting the appropriate scheme description.

Interactive analysis is limited to 500,000 data points (isolates x loci). If you select more than this then output will be restricted to static tables.

**Isolates**

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

Clear List all

**User genomes**

Optionally include data not in the database.

Upload assembly FASTA file (or zip file containing multiple FASTA files - one per genome):

Browse... No file selected.

**Loci**

- 16S\_rDNA
- 16S\_rRNA (SSU\_rRNA)
- 23S\_rRNA
- abcZ
- abcZ (NEIS1015)
- aceF (NEIS1279)
- ackA2 (NEIS1727)
- acnA (NEIS1729)

All None Paste list

**Recommended schemes**

Select one or more schemes below or use the full schemes list.

- MLST
- Ribosomal MLST
- N. meningitidis cgMLST v1.0
- N. gonorrhoeae cgMLST v1.0

Clear

**Schemes**

- eMLST (20 locus whc)
- N. gonorrhoeae cgML
- N. meningitidis cgMLST**
- OMV peptide typing
- Ribosomal MLST
- rplF species
- Other schemes

**Parameters / options**

Min % identity: 70

Min % alignment: 50

BLASTN word size: 20

**Action**

Reset Submit

The job will be sent to the job queue. When it has finished, you will have two options to display the output: ‘Pivot Table’ or ‘Heatmap’.

PubMLST Database home Contents

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

### Job status viewer

**Status**

Job id: BIGSdb\_194879\_1548675927\_08172

Submit time: 2019-01-28 11:45:27

Status: finished

Start time: 2019-01-28 11:45:43

Progress: 100%

Stop time: 2019-01-28 11:46:38

Total time: 54 seconds

**Output**

Pivot Table Heatmap

**Files**

Presence/absence (Excel)

Presence/absence (text)

Tar file containing all output files

Please note that job results will remain on the server for 7 days.

### 14.11.1 Pivot Table

Clicking the ‘Pivot Table’ button will display an interactive pivot table. The default display shows the number of isolates for which each locus is present or absent.

PubMLST Database home Contents

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

## Gene Presence - *Neisseria* PubMLST

158,895 data points (isolates x loci) used

Please note that table updates may be slow if you try to combine id (or isolate) with locus. Any other combination of fields is fine, e.g. locus vs presence, id vs presence.

**Pivot table** Change analysis

Drag and drop fields on to the table axes. Multiple fields can be combined.

Table id isolate complete known allele designated tagged

Count presence

locus

	presence	absent	present	Totals
locus				
NEIS0001	80	19	99	
NEIS0004	80	19	99	
NEIS0005	80	19	99	
NEIS0006	80	19	99	
NEIS0007	80	19	99	
NEIS0008	80	19	99	
NEIS0009	80	19	99	
NEIS0010	80	19	99	
NEIS0011	80	19	99	

You can break down any combination of fields by dragging them from the field area at the top of the table to either of the axes. For example, to show how many isolates have alleles designated and sequence regions tagged for each locus, drag the ‘designated’ and ‘tagged’ fields to the x-axis selector.

PubMLSTDatabase homeContents

Logged in: Keith Jolley (keith). Log out | Change passwordToggle: [icon]

Gene Presence - Neisseria PubMLST

158,895 data points (isolates x loci) used

Please note that table updates may be slow if you try to combine id (or isolate) with locus. Any other combination of fields is fine, e.g. locus vs presence, id vs presence.

Pivot table

Change analysis

Heatmap

Drag and drop fields on to the table axes. Multiple fields can be combined.

Table

id isolate complete known allele designated tagged

Count

presence

locus

	presence	absent	present	Totals
locus				
NEIS0001	80	19	99	
NEIS0004	80	19	99	
NEIS0005	80	19	99	
NEIS0006	80	19	99	
NEIS0007	80	19	99	
NEIS0008	80	19	99	
NEIS0009	80	19	99	
NEIS0010	80	19	99	
NEIS0011	80	19	99	

The table will be re-drawn including these fields.

Pivot table

Change analysis

Heatmap

Drag and drop fields on to the table axes. Multiple fields can be combined.

Table

id isolate complete known allele

Count

presence designated tagged

locus

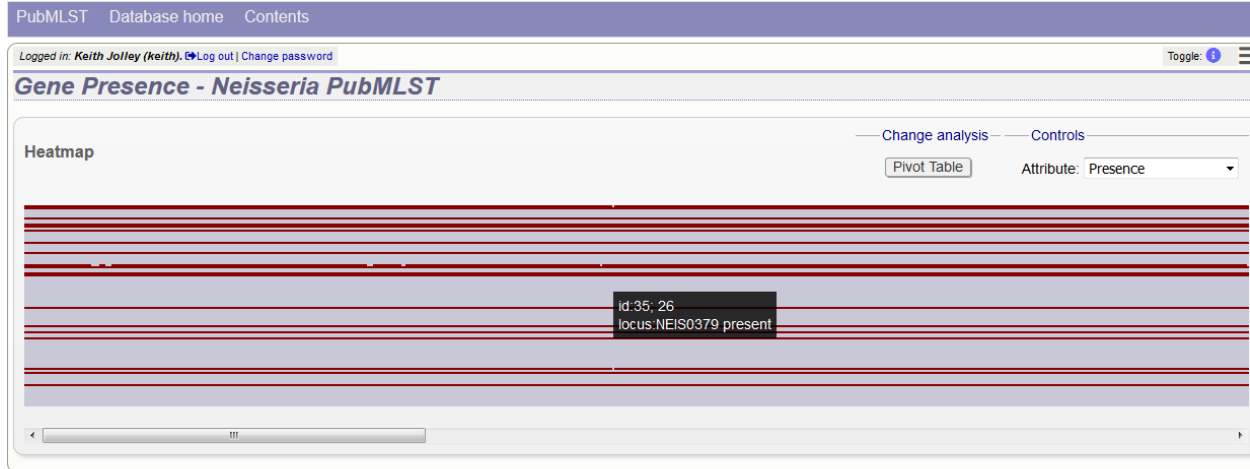
	presence	absent	present				Totals
	designated	not designated	designated		not designated		
	tagged	untagged	tagged	untagged	tagged	untagged	
locus							
NEIS0001		80	19				99
NEIS0004		80	19				99
NEIS0005		80	19				99
NEIS0006		80	19				99
NEIS0007		80	19				99
NEIS0008		80	19				99
NEIS0009		80	19				99
NEIS0010		80	19				99
NEIS0011		80	18		1		99
NEIS0012		80	19				99
NEIS0013		80	19				99
NEIS0014		80	19				99
NEIS0015		80	19				99
NEIS0016		80	17		2		99

**Note:** If your dataset has more than 100,000 data points (locus x isolates), then be aware that combining both id (or isolate) and locus within the table will result in sluggish performance. Any other combination of fields should be fine.

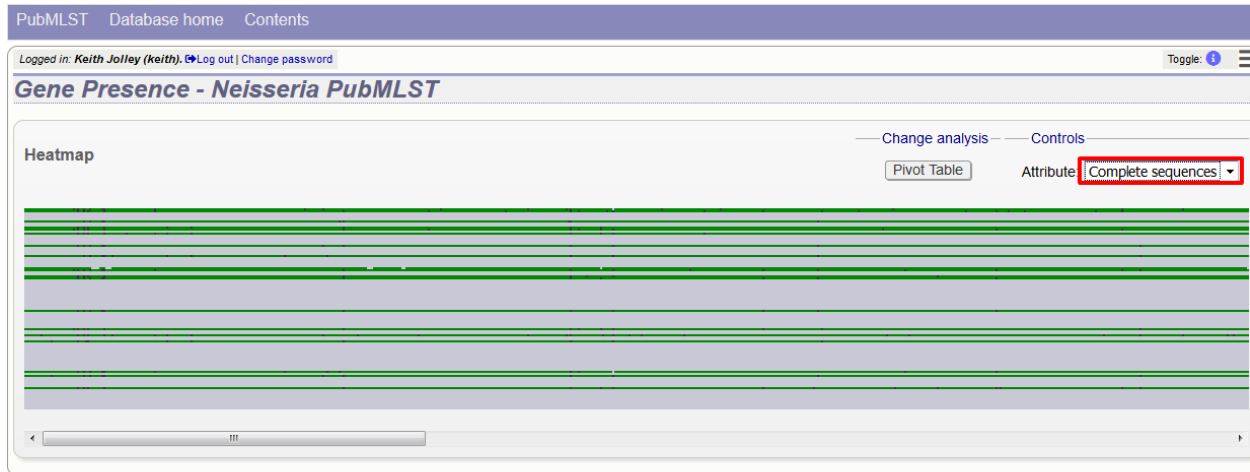
### 14.11.2 Heatmap

Clicking the ‘Heatmap’ button will display an interactive heatmap. By default the display shows the presence or absence of a locus for each isolate.

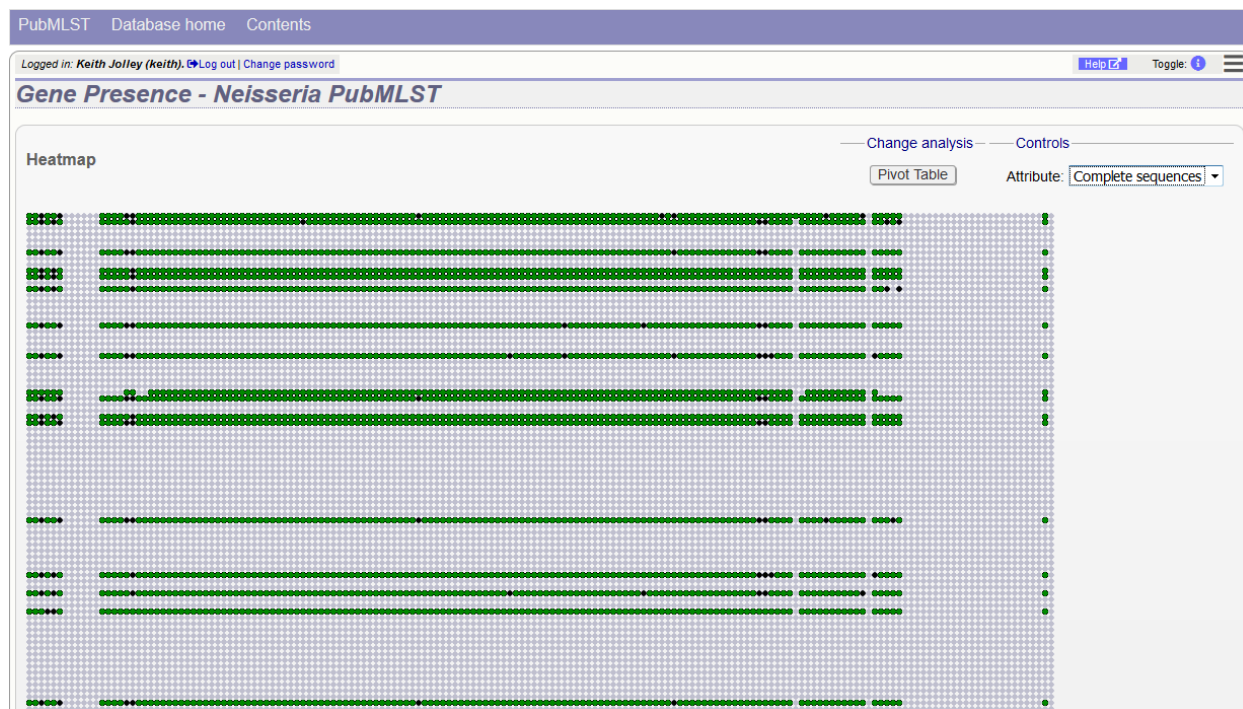
Hovering the mouse cursor or touching a region will identify the isolate and locus in a tooltip.



Change the attribute that is displayed by changing the selection in the attribute dropdown box:



The heatmap does scale to the number of records required to be displayed. If you find individual points to be too small, then choose a smaller subset of data to display:



## 14.12 GrapeTree

GrapeTree is a tool for generating and visualising minimum spanning trees. It has been developed to handle large datasets (in the region of 1000s of genomes) and works with 1000s of loci as used in cgMLST. It uses an improved minimum spanning algorithm that is better able to handle missing data than alternative algorithms and is able to produce publication quality outputs. Datasets can include metadata which allows nodes in the resultant tree to be coloured interactively.

GrapeTree can be accessed from the contents page by clicking the 'GrapeTree' link.

The screenshot shows the 'Neisseria PubMLST database' homepage. At the top, there is a navigation bar with 'PubMLST', 'Database home', and 'Contents'. Below this, a status bar indicates 'Log in'. The main title is 'Neisseria PubMLST database'. A descriptive paragraph states: '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.' The page is divided into several sections: 'Query database' (Search or browse database, Search by combinations of loci (profiles)), 'Projects' (Main public projects, Your projects), '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: 44,865, Last updated: 2018-02-20, Defined field values, 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), 'Miscellaneous' (Description of database fields), 'Analysis' (Codon usage, Presence/absence status of loci, Genome comparator, BLAST, Species identification), and 'Third party tools' (GrapeTree - Visualization of genomic relationships, iTOL - Phylogenetic trees with data overlays, Phylotree - Visualization and phylogenetic inference, Microreact - Open data visualization and sharing for genomic epidemiology). The 'GrapeTree' link is highlighted with a red box.

Alternatively, it can be accessed following a query by clicking the ‘GrapeTree’ button at the bottom of the results table. Isolates returned from the query will be automatically selected within the GrapeTree interface.

19023	M10 240473	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2170078	274	1191	ST-269 complex	7-2	4	F5-12
19024	M10 240480	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2267206	267	1194	ST-41/44 complex	7-2	4	F1-5
19025	M10 240481	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2194837	274	1831		22	9	F3-3
19026	M10 240482	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2184823	263	3754	ST-41/44 complex	7-2	4	F5-1
19027	M10 240484	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2168050	253	1097	ST-41/44 complex	17-1	23	F1-5
19028	M10 240485	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2193457	294	275	ST-269 complex	22	9	F5-12
19029	M10 240487	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2188678	241	1049	ST-269 complex	19-1	15-11	F5-1
19030	M10 240489	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2218969	275	461	ST-461 complex	19-2	13-1	F3-9
19031	M10 240490	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2175692	293	1161	ST-269 complex	22	9	F5-12
19032	M10 240498	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2154070	235	41	ST-41/44 complex	7-2	4	F1-5
19033	M10 240476	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2242902	284	9812	ST-213 complex	22	14	F5-5
19034	M10 240499	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2189734	198	275	ST-269 complex	22	9	F5-12
19035	M10 240500	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2197545	223	41	ST-41/44 complex	7-2	4	F1-5
19036	M10 240502	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2173298	188	340	ST-41/44 complex	7-2	4	F1-5
19037	M10 240503	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	Y		2101043	197	23	ST-23 complex	5-1	2-2	F1-96
19038	M10 240504	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	Y		2090713	189	1655	ST-23 complex	5-1	10-10	F4-1
19039	M10 240505	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	Y		2108738	170	183	ST-23 complex	21	16-5	deleted
19040	M10 240507	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2233910	196	1575		7-2	13-1	F1-7
19041	M10 240508	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2174583	219	4713		22	9	F5-12
19042	M10 240511	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2203751	205	41	ST-41/44 complex	7-1	1	F1-5
19043	M10 240512	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2110464	160	11	ST-11 complex	5	2	F1-1
19044	M10 240514	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	W		2181841	205	269	ST-269 complex	5-1	2-2	F5-1
19045	M10 240515	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	B		2095486	190	1655	ST-23 complex	5-1	10-1	F4-1
19046	M10 240520	UK	2010	invasive (unspecified/other)	Neisseria meningitidis	Y								

Analysis tools:

Breakdown: Fields Two Field Combinations Schemes/alleles Publications Sequence bin

Analysis: Codons Presence/Absence Genome Comparator BLAST rMLST species id

Export: Dataset Contigs Sequences

Third party: **GrapeTree** PhyloViz

Page: 1 2 3 4 5 6 7 8 9 > Last

Select the isolates to include. The tree can be generated from allelic profiles of any selection of loci, or more conveniently, you can select a scheme in the scheme selector to include all loci belonging to that scheme.

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

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

## GrapeTree: Visualization of genomic relationships - Neisseria PubMLST

This plugin generates a minimum-spanning tree and visualizes within GrapeTree:

GrapeTree: Visualization of core genomic relationships

GrapeTree is developed by:

- Zhemlin Zhou (1)
- Nabil-Fareed Alikhan (1)
- Martin J. Sergeant (1)
- Nina Luhmann (1)
- Cátia Vaz (2,5)
- Alexandre P. Francisco (2,4)
- João André Carriço (3)
- Mark Achtman (1)

1. Warwick Medical School, University of Warwick, UK
2. Instituto de Engenharia de Sistemas e Computadores: Investigação e Desenvolvimento (INESC-ID), Lisboa, Portugal
3. Universidade de Lisboa, Faculdade de Medicina, Instituto de Microbiologia and Instituto de Medicina Molecular, Lisboa, Portugal
4. Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
5. ADETC, Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Lisboa, Portugal

Publication: Zhou et al. (2017) GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens. [bioRxiv 216788](#).

This tool will generate minimum spanning trees from allelic profiles. Please check the loci that you would like to include. Alternatively select one or more schemes to include all loci that are members of the scheme. Analysis is limited to 50,000 records.

Isolates: 18968 18969 19023 19024 19025 19026 19027 19028

Loci: 16S\_rDNA 16S\_rRNA (SSU\_rRNA) 23S\_rRNA abcZ (NEIS1015) aceF (NEIS1279) ackA2 (NEIS1727) acnA (NEIS1729)

Schemes: 16S Antigen genes eMLST (20 locus partial genes) eMLST (20 locus whole genes) N gonorrhoeae cgMLST v1.0 **N. meningitidis cgMLST v1.0** Ribosomal MLST rplF species

Include fields: Select additional fields to include in GrapeTree metadata. country continent region year epidemiological year age yr

Action: Submit

The job will be sent to the job queue. When it has finished, click the button marked ‘Launch GrapeTree’.

PubMLST Database home Contents

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

### Job status viewer

**Status**

Job id: BIGSdb\_141866\_1519311802\_19416  
 Submit time: 2018-02-22 15:03:22  
 Status: finished  
 Start time: 2018-02-22 15:03:42  
 Progress: 100%  
 Stop time: 2018-02-22 15:07:38  
 Total time: 3 minutes and 55 seconds

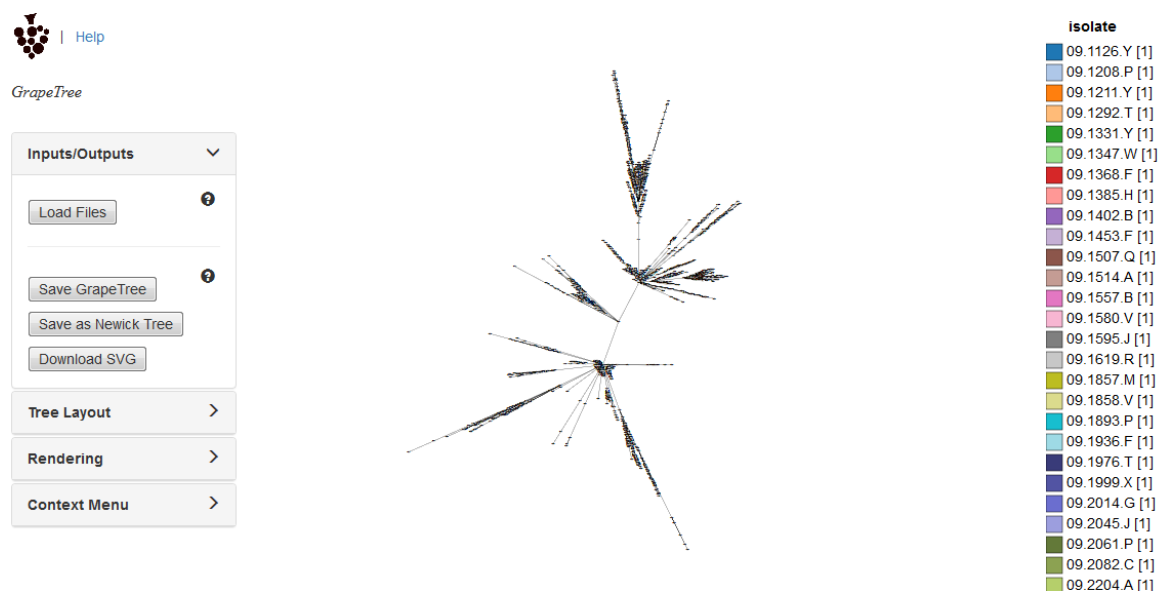
**Output**

[Launch GrapeTree](#)

- Profiles (TSV format) (14.2 MB)
- MS Tree (Newick format)
- Metadata (TSV format)
- Tar file containing output files (only files <10MB included - download larger files separately)

Please note that job results will remain on the server for 7 days.

The generated tree will be rendered in the GrapeTree application page.

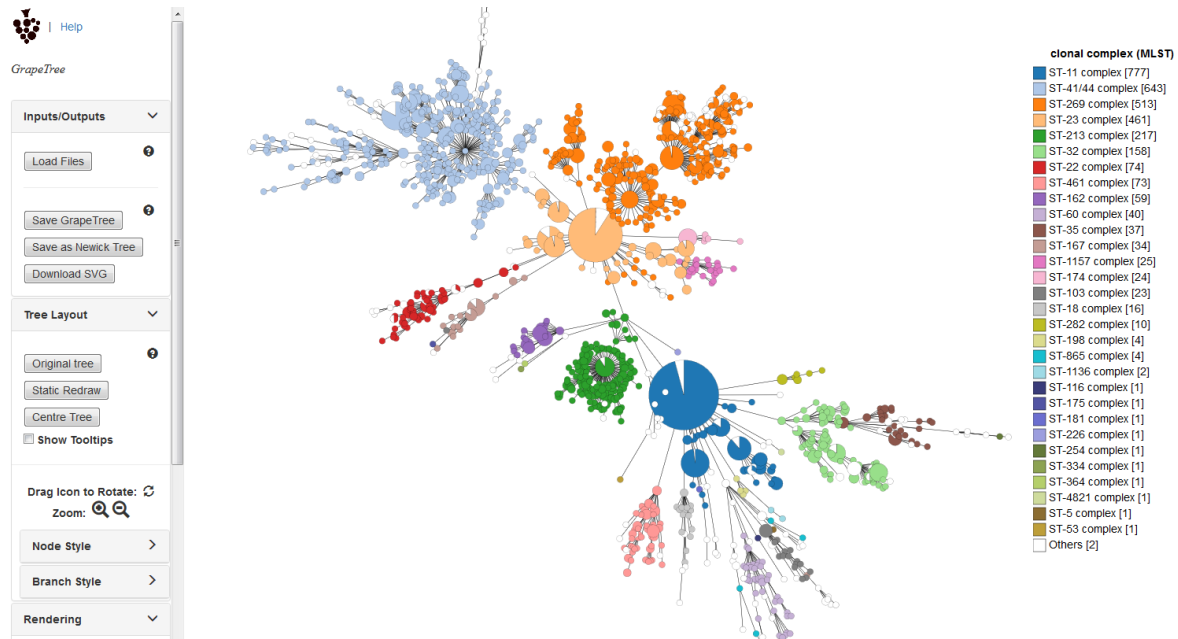


The image can be manipulated in various ways. These include modifying the tree layout, customising node labels and size, modifying branch lengths and collapsing branches. The image can be saved in SVG format which can be further edited in image publishing software such as Inkscape.

As an example, the default cgMLST tree (above) has been modified (below) as follows:

- Nodes coloured by clonal complex
- Labels removed
- Branches collapsed where  $\leq 100$  loci different
- Node size set to 200%
- Kurtosis (node size relative to number of isolates) set to 75%
- Dynamic rendering allowed to run to fan out nodes





Full details can be found in the [GrapeTree manual](#).

**Note:** GrapeTree has been described in the following publication:

Z Zhou, NF Alikhan, MJ Sergeant, N Luhmann, C Vaz, AP Francisco, JA Carrico, M Achtman (2018) GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens. [Genome Res 28:1395-1404](#).



## 15.1 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.

The screenshot shows the Neisseria PubMLST database interface. At the top, there 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. Below this is the title 'Neisseria PubMLST database'. A descriptive paragraph states: '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.'

The main content area is divided into several sections:

- Query database** (magnifying glass icon):
  - 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** (gear icon):
  - Set general options - including isolate table field handling.
  - Set display and query options for locus, schemes or scheme fields.
- Submissions** (upload icon):
  - Manage submissions
- General information** (info icon):
  - Isolates: 34218
  - Last updated: 2015-06-30
  - Update history
  - About BIGSdb

At the bottom, there is a navigation bar with four main sections:

- Breakdown** (pie chart icon):
  - Single field
  - Two field
  - Unique combinations
  - Scheme and alleles
  - Publications
  - Sequence bin
- Export** (floppy disk icon):
  - Export dataset** (highlighted with a red box)
  - Contigs
  - Sequences - XMFA / concatenated FASTA formats
- Analysis** (line graph icon):
  - Codon usage
  - Presence/absence status of loci
  - Genome comparator
  - BLAST
- Miscellaneous** (document icon):
  - 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								MLST		Finotyping antigens		
id	isolate	aliases	country	year	disease	species	serogroup	ST	clonal complex	PorA VR1	PorA VR2	FoTA 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	M00242027		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

☐ 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

☐ sender

☐ curator

☐ date entered

☐ datestamp

**Composite fields**

☐ strain designation

**References**

☐ references

☐ PubMed id

☐ Full citation

**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

☒ 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

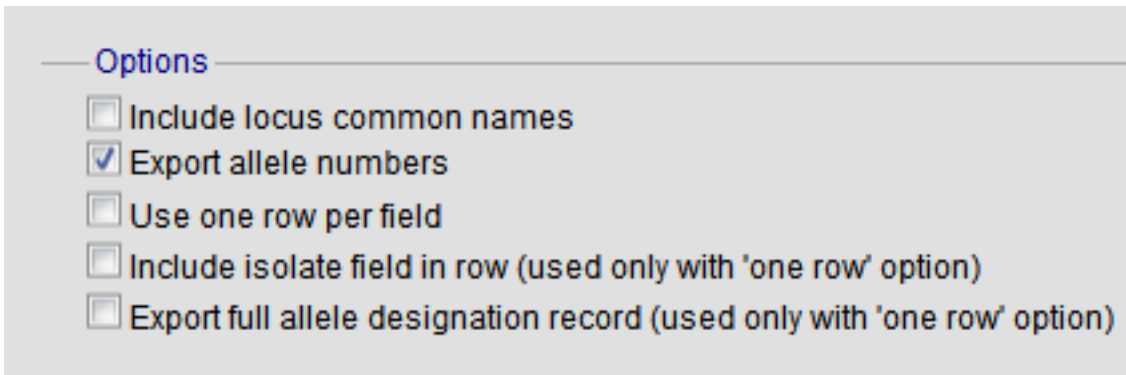
☒ GTG/TTG at start codes for methionine

Click Submit.

You can then download the data in tab-delimited text or Excel formats.



### 15.1.1 Advanced options

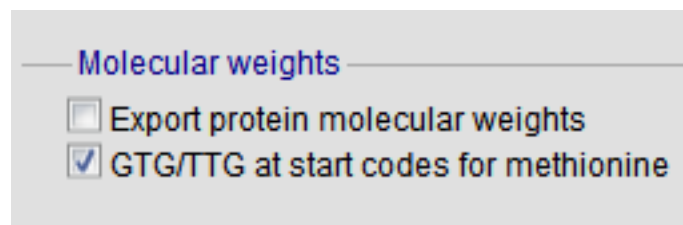


The options fieldset has the following options.

- 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' format.
- Export full allele designation record - export sender, curator and datestamp information as separate rows when exporting allele designation data.

### 15.1.2 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.

## 15.2 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	M102/	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 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	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.

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

Toggle: [?](#)

### 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) acfF (NEIS1729) ackA2 (NEIS1729) acnA (NEIS1729)

Schemes:  Genetic information Prote Metabolism Pili 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.

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

Toggle: [?](#)

### 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.



The screenshot shows the 'Job status viewer' interface. At the top, there 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, and Database submissions. The main content area is titled 'Job status viewer' and contains a 'Status' section with the following details:

- 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

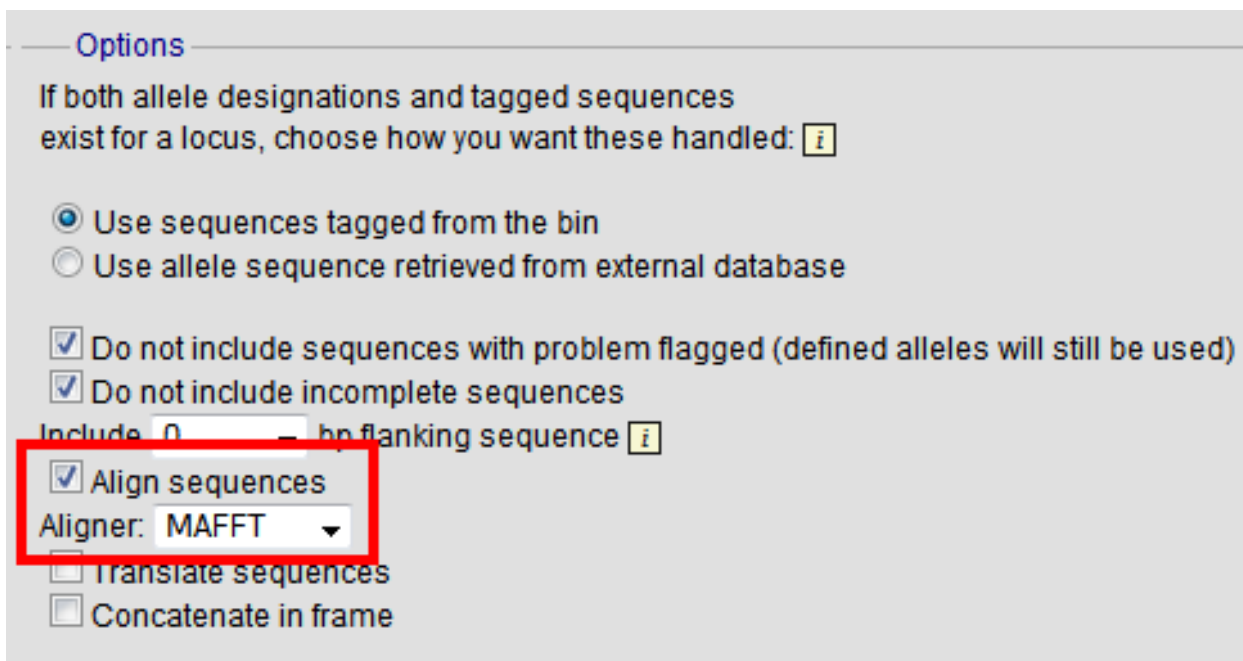
Below the status section is an 'Output' section with a red box highlighting the following items:

- XMFA output file (not aligned)
- Concatenated FASTA (not aligned)
- Tar file containing output files

At the bottom, a green box contains the text: 'Please note that job results will remain on the server for 7 days.'

### 15.2.1 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.



The screenshot shows the 'Options' section of the interface. It contains the following settings:

- 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

The 'Align sequences' checkbox and the 'Aligner' dropdown menu are highlighted with a red box.

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.



## 15.3 Profile export

You can export the allelic profiles for any indexed scheme (those containing a primary key field) defined in the sequence definition database.

The profile export function can be accessed by clicking the ‘Profiles’ link in the Export section of the contents page.

PubMLST Database home Contents

Log in Toggle

### Neisseria profile/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 or genome.
- Batch sequence query - query multiple sequences in FASTA format.
- Sequence attribute search - find alleles by matching criteria (all loci together)
- Locus-specific sequence attribute search - select, analyse and download specific alleles.
- Search, browse or enter list of profiles
- Search by combinations of alleles - including partial matching.
- Batch profile query - lookup profiles copied from a spreadsheet.
- Extract finetype from whole genome data

#### Downloads

- Allele sequences
- MLST

#### Option settings

- Set general options
- Scheme options

#### Submissions

- Manage submissions

#### General information

- Number of sequences: 1135267
- Number of profiles: Show
- Last updated: 2018-01-24
- Profile update history
- About BIGSdb

#### Export

- Profiles**
- 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.

Alternatively, you can access this function by clicking the ‘Profiles’ button in the Export list at the bottom of a results table. Please note that the list of functions here may vary depending on the setup of the database.

Page: 1 2 3 4 5 6 7 8 9 > Last

ST	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	clonal complex
1	1	3	1	1	1	1	3	ST-1 complex/subgroup VII
2	1	3	4	7	1	1	3	ST-1 complex/subgroup VII
3	1	3	1	1	1	23	13	ST-1 complex/subgroup VII
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
15	13	3	16	1	3	11	9	ST-364 complex
16	15	9	9	13	8	19	15	ST-424/6688 complex
17	8	3	13	1	11	12	4	
18	7	8	10	19	10	1	2	ST-18 complex
19	7	8	10	19	8	1	2	ST-18 complex
20	6	8	10	17	10	1	2	ST-18 complex
21	1	5	1	1	2	16	17	
22	11	5	18	8	11	24	21	ST-22 complex
23	10	5	18	9	11	9	17	ST-23 complex/Cluster A3
24	2	5	2	7	15	20	5	ST-750 complex
25	6	5	2	12	6	13	14	

Analysis tools:

Export: **Profiles** Sequences

Page: 1 2 3 4 5 6 7 8 9 > Last

This will take you to a form with a list box in which the identifiers of the profile definitions you wish to include can be entered. Following a query, these values will be pre-entered. If the box is left empty then all profiles will be included.

PubMLST Database home Contents

Log in

### Export allelic profiles - *Neisseria* profile/sequence definitions

**Schemes**

Please select the scheme you would like to query:

MLST

This script will export allelic profiles in tab-delimited text and Excel formats.

**Select STs** **Action**

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

Click submit.

The export job will be submitted to the job queue.

PubMLST Database home Contents

Log in

### Job status viewer

**Status**

Job id: BIGSdb\_012009\_1516786630\_99221

Submit time: 2018-01-24 09:37:10

Status: finished

Start time: 2018-01-24 09:37:24

Progress: 100%

Stop time: 2018-01-24 09:37:36

Total time: 12 seconds

**Output**

- Profiles - Tab-delimited text [Download](#)
- Profiles - Excel format [Download](#)
- Tar file containing output files [Download](#)

Please note that job results will remain on the server for 7 days.

The profiles will be exported in tab-delimited text and Excel formats.

## 15.4 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)	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)	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		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:

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

AllNone

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

AllNone

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.

350

Chapter 15. Data export plugins

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)

### 15.4.1 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.


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](#) (experimental) **NEW**




### Downloads

- [Allele sequences](#)
- [MLST](#)




### Option settings

- [Set general options](#)




### Submissions

- [Manage submissions](#)




### General information

- Number of sequences: 124772
- Number of profiles: [Show](#)
- 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.

## 16.1 Registering a user account

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


## 16.2 Allele submission

New allele data can only be submitted from within the appropriate sequence definition database. Submissions consist of one or more new allele sequences for a single locus. You will need to create separate submissions for each locus - this is because different loci may be handled by different curators.

### 16.2.1 Start

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





[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](#)

### 16.2.2 Select the submission locus

Select the locus from the locus list box:



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

Logged in: [Keith Jolley \(keith2\)](#) | [Log out](#) | [Change password](#)

[Help](#)

## Manage submissions

### Submit new 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:

read length:

coverage:

assembly:

assembly software:

**FASTA or single sequence**

**Action**

The locus list may be very long in some databases. It may be possible to filter these to those belonging to specific schemes. If the scheme tree is shown, select the appropriate scheme, e.g. 'MLST' and click 'Filter'.

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

**Select locus**

Filter

- abcZ
- abcZ (NEIS1015)
- aceF (NEIS1279)
- acnA (NEIS1729)
- acnB (NEIS1492)
- adk
- adk (NEIS0767)
- 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

**Select locus**

Filter

- abcZ
- adk
- aroE
- fumC
- gdh
- pdhC
- pgm

**Sequence details**

technology: Illumina  
read length: 100-199  
coverage: 20-49x  
assembly: de novo  
assembly software: Velvet

**FASTA or single sequence**

Action  
Submit

### 16.2.3 Enter details of sequencing method

There are a number of fields that must be filled in so that the curator knows how the sequence was obtained:

- technology - the sequencing platform used, allowed values are:

- 454
  - Illumina
  - Ion Torrent
  - PacBio
  - Oxford Nanopore
  - Sanger
  - Solexa
  - SOLiD
  - other
  - unknown
- read length - this is the length of sequencing reads. This is a required field for Illumina data, and not relevant to Sanger sequencing. Allowed values are:
  - <100
  - 100-199
  - 200-299
  - 300-499
  - >500
- coverage - the mean number of reads covering each nucleotide position of the sequence. This is not relevant to Sanger sequencing. Allowed values are:
  - <20x
  - 20-49x
  - 50-99x
  - >100x
- assembly - the means of generating the submitted sequence from the sequencing reads. Allowed values are:
  - de novo
  - mapped
- assembly software - this is a free text field where you should enter the name of the software used to generate the submitted sequence.

#### 16.2.4 Paste in sequence(s)

Paste in the new variant sequences to the box. This can either be a stand- alone sequence or multiple sequences in FASTA format. The sequences must be trimmed to the start and end points of the loci - check existing allele definitions if in doubt. The submission is likely to be rejected if sequences are not trimmed. Click submit.

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
  - Finetyping antigens
  - Antigen genes
  - eMLST (20 locus partial genes)
  - eMLST (20 locus whole genes)
  - Other schemes
  - Loci not in schemes

Filter

Select locus: abcZ

Sequence details

technology: Illumina  
read length: 100-199  
coverage: 20-49x  
assembly: de novo  
assembly software: Velvet

FASTA or single sequence

```
>NM322
TTTGATACTGTCGCCGAAGGTTTGGCGAAATTCGCGAATTAT
TGCGCCGTTATCATCATGTGAGCCATGAGTTGAAATGGTTC
GAGTGAGGCTTATTGAAAGAGCTCAACGAATTGCAACTTGAG
ATCGAAGCGAAGGACGGCTGGAAGTTGGATGCGCGGTGAAGC
AGACTTTGGCGAAGCTCGGTTTGGCGAAACGAAAAATCGG
CAACCTCTCGGCGATCAGAAAAAGCGGTCGCTTGGCGGAG
```

Action

Submit

The system will perform some basic checks on the submitted sequences. If any of the sequences have been defined previously they must be removed from the submission before you can proceed. Curators do not want to waste their time dealing with previously defined sequences.

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

**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
  - Finetyping antigens
  - Antigen genes
  - eMLST (20 locus partial genes)
  - eMLST (20 locus whole genes)
  - Other schemes
  - Loci not in schemes

Filter

Select locus: abcZ

Sequence details

technology: Illumina  
read length: 100-199  
coverage: 20-49x  
assembly: de novo  
assembly software: Velvet


FASTA or single sequence

```
>NM322
TTTGATACTGTCGCCGAAGGTTTGGCGAA
ATTGCGAATTATTGCGCGTTATCATCAT
GTCAGCCATGAGTTGAAATGGTTGAGT
GAGGCTTATTGAAAGAGCTCAACGAATTG
CAACTTGAGATCGAAGCGAAGGACGGCTG
```

Action

Submit

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 \(keith2\)](#) | [Log out](#) | [Change password](#)
[Help](#)

## Manage submissions

Submission: BIGSdb\_20150709114223\_7557\_43592

[Sequences](#)

You are submitting the following abcZ sequences: [Download](#) 

Identifier	Length	Sequence	Status	Assigned allele
NM322	433	TTTGATACTGTTGCCGAACG ... 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 [kajolley@gmail.com](mailto:kajolley@gmail.com).

☒ 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](#)

Message: [Append](#)

### 16.2.5 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Append'.


**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 (keith2) Log out | Change password Help

## Manage submissions

Submission: BIGSdb\_20150709114223\_7557\_43592

**Sequences**

You are submitting the following abcZ sequences: [Download](#) 

Identifier	Length	Sequence	Status	Assigned allele
NM322	433	TTTGATACTGTTGCCGAACG ... 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 kajolley@gmail.com.

☒ 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.

Message: [Append](#)

The message will be attached. A curator may respond to the message and attach their own, with the full conversation becoming part of the submission record.



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

Logged in: [Keith Jolley \(keith2\)](#) | [Log out](#) | [Change password](#)
[Help](#)

## Manage submissions

Submission: BIGSdb\_20150709114223\_7557\_43592

**Sequences**

You are submitting the following abcZ sequences: [Download](#) 

Identifier	Length	Sequence	Status	Assigned allele
NM322	433	TTTGATACTGTTGCCGAACG ... GCGGATTGTCGAACTCGATC	pending	
NM21	433	TTTGATACCGTTGCCGAAGG ... GCGGATTGTCGAACTCGATC	pending	

**Sequence details**

technology:

read length:

coverage:

assembly:

assembly software:

**E-mail**

Updates will be sent to [kajolley@gmail.com](mailto:kajolley@gmail.com).

☒ 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
2015-07-09 10:44:46+00	Keith Jolley	The sequence variant NM322 has been seen in 3 isolates.

Message: [Append](#)

### 16.2.6 Add supporting files

Some submissions will require the attachment of supporting files. This will depend on the policies of the individual databases. Sequences determined by Sanger sequencing should normally have forward and reverse trace files attached. Files can be added to the submission by clicking the 'Browse' button in the 'Supporting files' section.



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

Logged in: [Keith Jolley \(keith2\)](#) | [Log out](#) | [Change password](#)
[Help](#)

## Manage submissions

Submission: BIGSdb\_20150709114223\_7557\_43592

### Sequences

You are submitting the following abcZ sequences: [Download](#) 

Identifier	Length	Sequence	Status	Assigned allele
NM322	433	TTTGATACTGTTGCCGAACG ... GCGGATTGTCGAACTCGATC	pending	
NM21	433	TTTGATACCGTTGCCGAAGG ... GCGGATTGTCGAACTCGATC	pending	

### Sequence details

technology:

read length:

coverage:

assembly:

assembly software:

### E-mail

Updates will be sent to [kajolley@gmail.com](mailto:kajolley@gmail.com).

☒ 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.

No files selected.


### Messages

Timestamp	User	Message
2015-07-09 10:44:46+00	Keith Jolley	The sequence variant NM322 has been seen in 3 isolates.

Message:

Select the file in the selection box, then click ‘Upload files’.





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

Logged in: **Keith Jolley (keith2)** | [Log out](#) | [Change password](#)
[Help](#)

## Manage submissions

Submission: BIGSdb\_20150709114223\_7557\_43592

### Sequences

You are submitting the following abcZ sequences: [Download](#) 

Identifier	Length	Sequence	Status	Assigned allele
NM322	433	TTTGATACTGTTGCCGAACG ... GCGGATTGTCGAACTCGATC	pending	
NM21	433	TTTGATAACCGTTGCCGAAGG ... GCGGATTGTCGAACTCGATC	pending	

### Sequence details

technology:   
read length:   
coverage:   
assembly:   
assembly software:

### E-mail

Updates will be sent to [kajolley@gmail.com](mailto:kajolley@gmail.com).

☒ 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.

### Messages

Timestamp	User	Message
2015-07-09 10:44:46+00	Keith Jolley	The sequence variant NM322 has been seen in 3 isolates.

Message: [Append](#)

The file will be uploaded and shown in a 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 (keith2) Log out | Change password Help

## Manage submissions

Submission: BIGSdb\_20150709114223\_7557\_43592

Sequences

You are submitting the following abcZ sequences: [Download](#) 

Identifier	Length	Sequence	Status	Assigned allele
NM322	433	TTTGATACGTGCGGAACG ... 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 kajolley@gmail.com.

☒ 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
2015-07-09 10:44:46+00	Keith Jolley	The sequence variant NM322 has been seen in 3 isolates.

Message: [Append](#)

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

### 16.2.7 Finalize submission

Make sure the 'E-mail submission updates' box is checked if you wish to receive E-mail notification of the result of your submission. This setting is remembered between submissions.

Click 'Finalize submission!'.

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 (keith2) Log out | Change password Help

### Manage submissions

Submission: BIGSdb\_20150709114223\_7557\_43592

**Sequences**

You are submitting the following abcZ sequences: [Download](#)

Identifier	Length	Sequence	Status	Assigned allele
NM322	433	TTTGATACGTGTGCCGACG ... GCGGATTGTCGAACTCGATC	pending	
NM21	433	TTTGATACGTGTGCCGACG ... 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 kajolley@gmail.com.

☒ 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	<a href="#">Delete</a>

[Delete selected files](#)

**Messages**

Timestamp	User	Message
2015-07-09 10:44:46+00	Keith Jolley	The sequence variant NM322 has been seen in 3 isolates.

Message:  [Append](#)

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

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 (keith2) Log out | Change password Help

### Manage submissions

**Submit new data**

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

**Submission type:**

- alleles
- MLST profiles

**Pending submissions**

You have submitted the following submissions that are pending curation:

Submission id	Submitted	Updated	Type	Details
BIGSdb_20150709114223_7557_43592	2015-07-09	2015-07-09	alleles	2 abcZ sequences

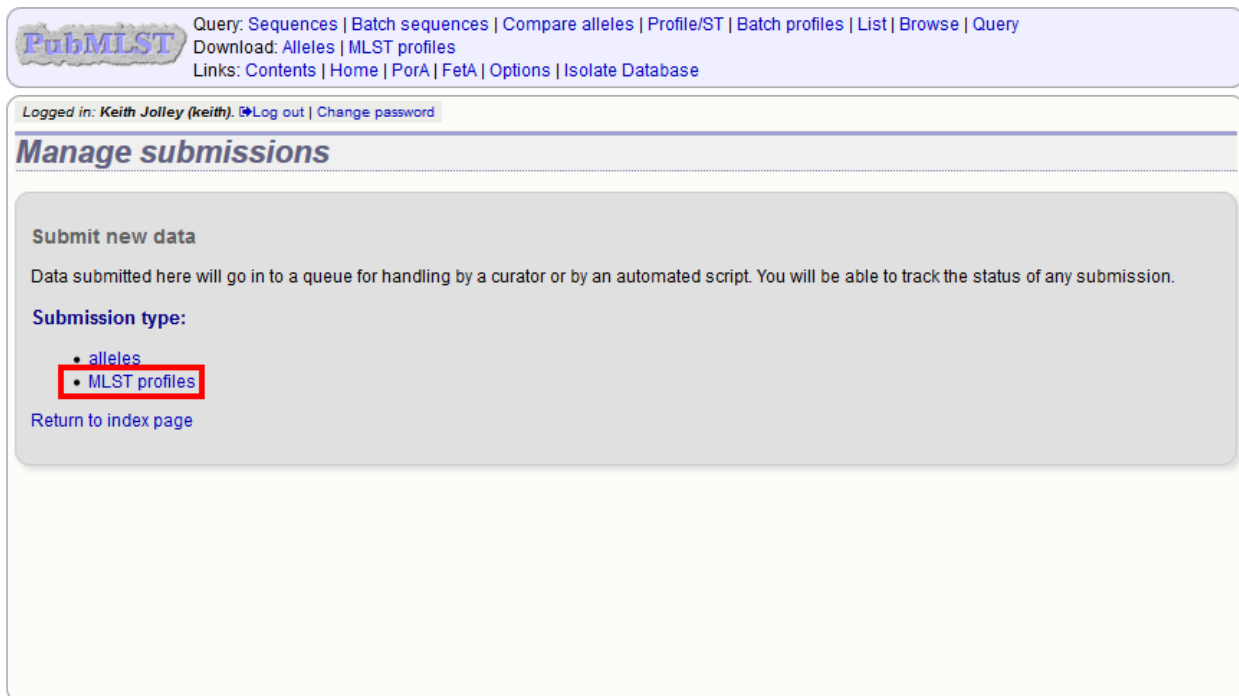
[Return to index page](#)

## 16.3 Profile submission

### 16.3.1 Start

**Note:** Most MLST databases on PubMLST.org require you to submit an isolate record for each new ST that you wish to be defined. In these cases, you should add the isolate name to the id field of your profile submission and make a corresponding *isolate submission* containing the allelic profile.

Click the appropriate profiles link under submission type on the submission management page.



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**

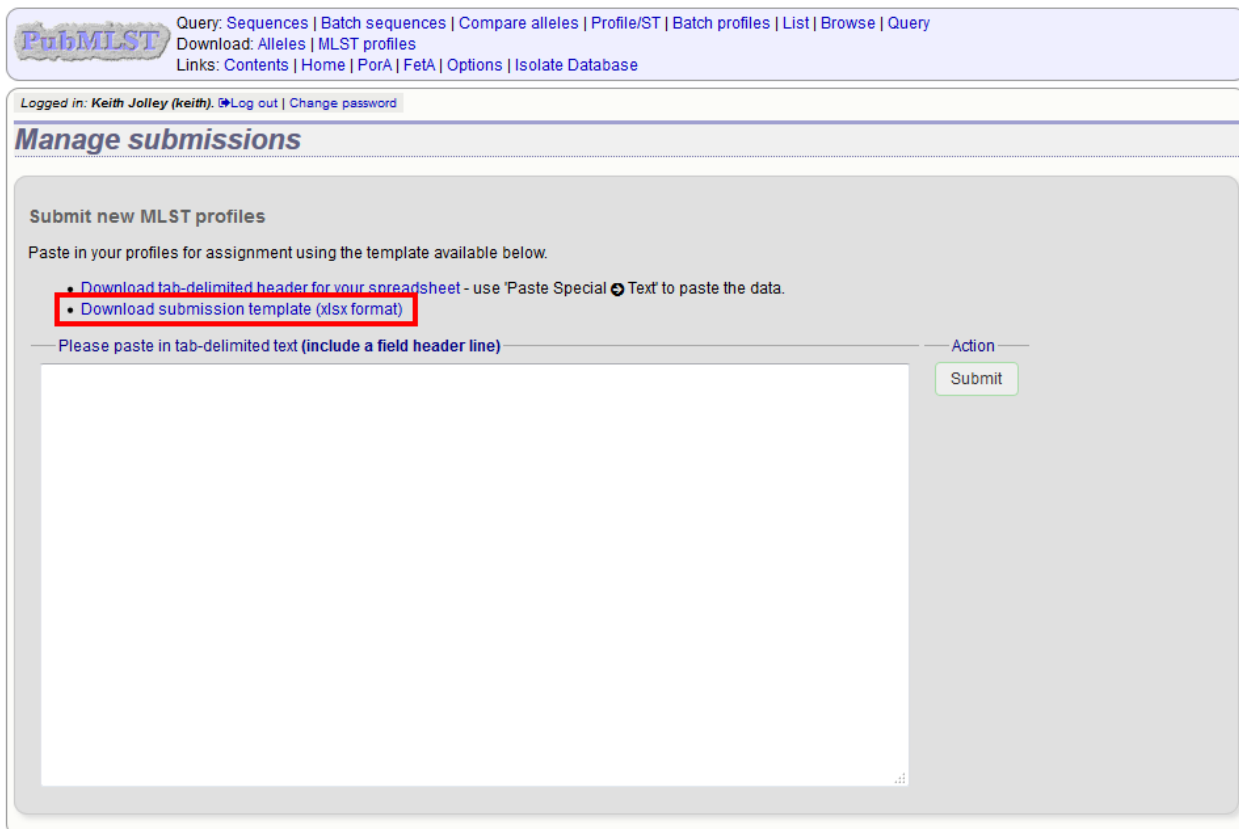
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

**Submission type:**

- alleles
- MLST profiles**

[Return to index page](#)

Download the Excel submission template.



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 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

### 16.3.2 Paste in profile(s)

Fill in the template. The first column 'id' can be used to enter an identifier that is meaningful to you - it is used to report back the results but is not uploaded to the database. It can be left blank, or the entire column can be removed - in which case individual profiles will be identified by row number.

Copy and paste the entire contents of the submission worksheet. Click submit.


[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	3	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

### 16.3.3 Add message to curator


If you wish to enter a message to the curator, enter this in the messages box and click 'Append'.

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 (keith2) Log out | Change password Help

### Manage submissions

Submission: BIGSdb\_20150709115437\_16625\_11378

**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 kajolley@gmail.com.  
☒ 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**  
 Corresponding isolate data has already been submitted (ids 43721-43722).  
 Message: Append


The message will be attached. A curator may respond to the message and attach their own, with the full conversation becoming part of the submission record.

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 (keith2) Log out | Change password Help

### Manage submissions

Submission: BIGSdb\_20150709115437\_16625\_11378

**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 kajolley@gmail.com.  
☒ 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
2015-07-09 10:56:00+00	Keith Jolley	Corresponding isolate data has already been submitted (ids 43721-43722).

Message: Append

### 16.3.4 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'.

### 16.3.5 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!’.

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 (keith2) Log out | Change password Help

### Manage submissions

Submission: BIGSdb\_20150709115437\_16625\_11378

**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 kajolley@gmail.com.  
☒ 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
2015-07-09 10:56:00+00	Keith Jolley	Corresponding isolate data has already been submitted (ids 43721-43722).

Message: [Append](#)

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 (keith2) Log out | Change password Help

### Manage submissions

**Submit new data**  
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

**Submission type:**

- alleles
- MLST profiles

**Pending submissions**  
You have submitted the following submissions that are pending curation:

Submission id	Submitted	Updated	Type	Details
BIGSdb_20150709115437_16625_11378	2015-07-09	2015-07-09	profiles	2 MLST profiles

[Return to index page](#)

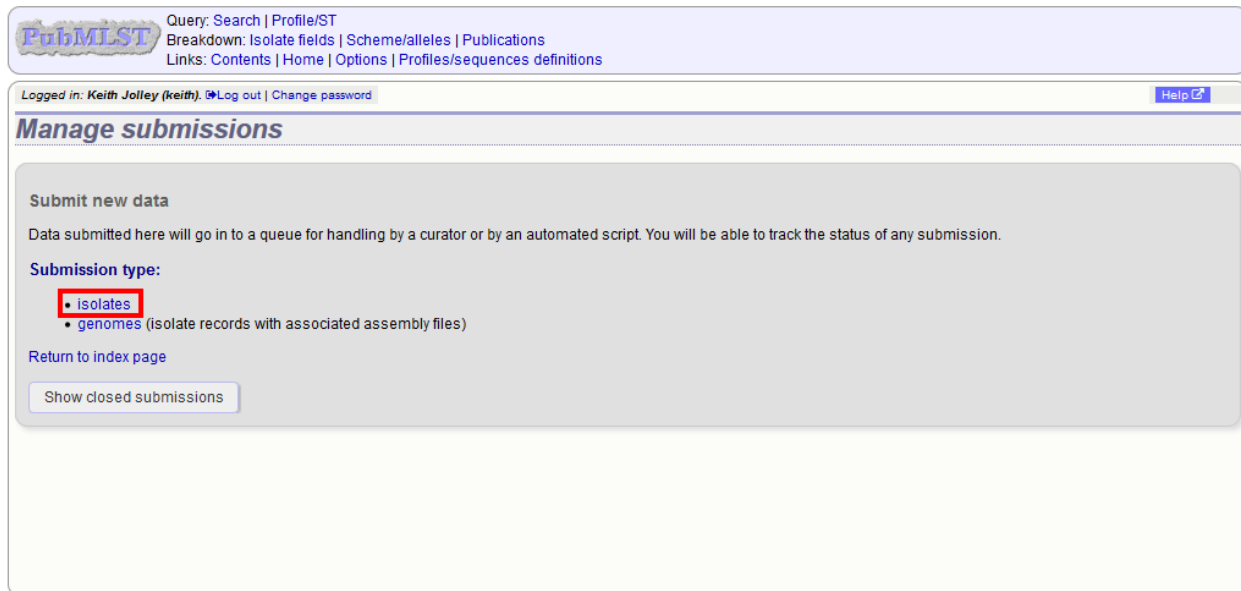


## 16.4 Isolate submission

New isolate data can only be submitted from within the appropriate isolate database. You may be required to submit isolate data if you would like to get a new MLST sequence type defined, but this depends on individual database policy.

### 16.4.1 Start

Click the ‘isolates’ link under submission type on the submission management page.



PubMLST Query: Search | Profile/ST  
Breakdown: Isolate fields | Scheme/alleles | Publications  
Links: Contents | Home | Options | Profiles/sequences definitions

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

### Manage submissions

**Submit new data**  
Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.


**Submission type:**

- isolates**
- genomes (isolate records with associated assembly files)

[Return to index page](#)

Show closed submissions

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

### 16.4.2 Paste in isolate data

Fill in the template. Some fields are required and cannot be left blank. Check the 'Description of database fields' link on the database contents page to see a description of the fields and allowed values where these have been defined. Where allowed values have been set, the template will have dropdown boxes (although these require newer versions of Excel to work).

Some databases may have hundreds of loci defined, and most will not have a column in the template. You can add new columns for any loci that have been defined and for which you would like to include allelic information for. These locus names must be the primary locus identifier. A list of loci can be found in the 'allowed\_loci' tab of the Excel submission template.

Copy and paste the entire contents of the submission worksheet. Click submit.

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 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		range						
rifampicin		rifampicin	range			ciprofloxacin		range						
pending assembly		comments				abc2		adk				aroE		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					

Action  
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 meningitis and septicaemia blood 2015 Neisseria meningitidis  
B  
2 3 4 3 8 4 6 F1-5 5 2  
UK325 septicaemia CSF UK 2015 Neisseria meningitidis  
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.

### 16.4.3 Add message to curator

If you wish to enter a message to the curator, enter this in the messages box and click 'Append'.

Submission: BIGSdb\_20150709121747\_1342\_99624

**Isolates**

You are submitting the following isolates: [Download](#) 

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

**E-mail** **Action**

Updates will be sent to keith.jolley@zoo.ox.ac.uk. [Finalize submission!](#)

☒ 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**

I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.

Message: [Append](#)

The message will be attached. A curator may respond to the message and attach their own, with the full conversation becoming part of the submission record.


**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](#) [Help](#) 

**Manage submissions**

Submission: BIGSdb\_20150709121747\_1342\_99624

**Isolates**

You are submitting the following isolates: [Download](#) 

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

**E-mail** **Action**

Updates will be sent to keith.jolley@zoo.ox.ac.uk. [Finalize submission!](#)

☒ 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
2015-07-09 11:19:12+00	Keith Jolley	I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.

Message: [Append](#)

#### 16.4.4 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.

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 Help

### Manage submissions

Submission: BIGSdb\_20150709121747\_1342\_99624

Isolates

You are submitting the following isolates: Download

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

E-mail

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
2015-07-09 11:19:12+00	Keith Jolley	I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.

Message: Append

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 Help

### Manage submissions

Submission: BIGSdb\_20150709121747\_1342\_99624

Isolates

You are submitting the following isolates: Download

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

E-mail

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... UK325\_genome\_contigs.fas Upload files

Messages

Timestamp	User	Message
2015-07-09 11:19:12+00	Keith Jolley	I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.

Message: Append

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](#) [Help](#)

## Manage submissions

Submission: BIGSdb\_20150709121747\_1342\_99624

Isolates

You are submitting the following isolates: [Download](#) 

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	tumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

E-mail

Updates will be sent to [keith.jolley@zoo.ox.ac.uk](mailto: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
2015-07-09 11:19:12+00	Keith Jolley	I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.

[Message:](#) [Append](#)

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

### 16.4.5 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 Help

## Manage submissions

Submission: BIGSdb\_20150709121747\_1342\_99624

Isolates

You are submitting the following isolates: [Download](#) 

isolate	country	disease	source	species	serogroup	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	FetA_VR	PorA_VR1	PorA_VR2
UK322	UK	meningitis and septicaemia	blood	Neisseria meningitidis	B	2	3	4	3	8	4	6	F1-5	5	2
UK325	UK	septicaemia	CSF	Neisseria meningitidis	B	2	3	4	3	18	4	6	F1-5	5-1	2

E-mail Action

Updates will be sent to keith.jolley@zoo.ox.ac.uk. [Finalize submission!](#)

☒ 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
2015-07-09 11:19:12+00	Keith Jolley	I think UK325 has a new MLST profile requiring definition of a new ST number. Thanks.

Message: [Append](#)

Your submission will then be listed under ‘Pending submissions’ on your submission 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 Help

## Manage submissions

Submit new data

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

Submission type:

- isolates

Pending submissions

You have submitted the following submissions that are pending curation:

Submission id	Submitted	Updated	Type	Details
BIGSdb_20150709121747_1342_99624	2015-07-09	2015-07-09	isolates	2 isolates

[Return to index page](#)

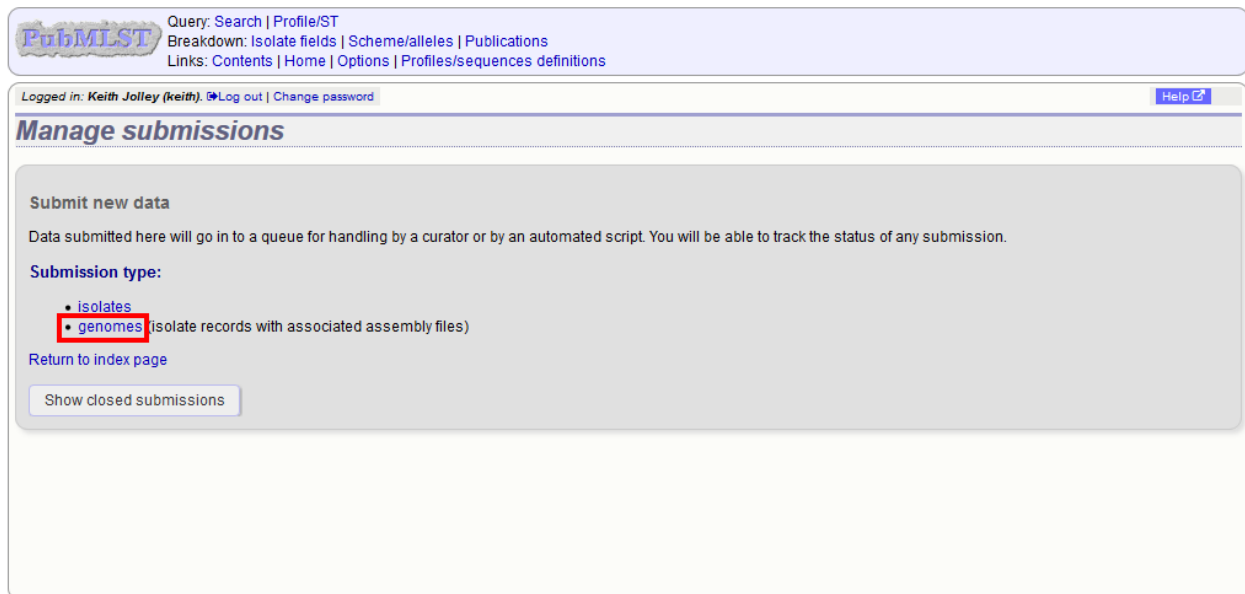


## 16.5 Genome submission

Submitting genomes uses the same process as standard *isolate submission*. The only difference is that there are a couple of extra required fields in the submission table:

- `assembly_filename` - this is the name of the FASTA file containing the assembly contigs. This must be uploaded as a supporting file - you will not be able to finalize the submission until every isolate record has a matching contig file.
- `sequence_method` - the sequencing technology used to generate the sequences. The allowed values are listed on the submission page.

To start the submission, click the ‘genomes’ link under submission type on the submission management page.



Then follow the steps for *isolate submission*, uploading the contigs files as supporting files.

## 16.6 Removing submissions from your notification list

Once a submission has been closed by a curator, the results will be displayed in your ‘Manage submissions’ area. You can remove submissions once you have noted the result by clicking the ‘Remove’ link.


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

Logged in: [Keith Jolley \(keith2\)](#) | [Log out](#) | [Change password](#)
[Help](#)

## Manage submissions

**Submit new data**

Data submitted here will go in to a queue for handling by a curator or by an automated script. You will be able to track the status of any submission.

**Submission type:**

- [alleles](#)
- [MLST profiles](#)

**Recently closed submissions**

You have submitted the following submissions which are now closed:

Submission id	Submitted	Updated	Type	Details	Outcome	Remove
BIGSdb_20150714071515_10601_27668	2015-07-14	2015-07-14	alleles	3 NEIS0001 (lpxC) sequences		

[Return to index page](#)

Alternatively, submissions will be removed automatically a specified period of time after closure. By default, this time is 90 days, but this can vary depending on the site configuration.

---

## RESTful Application Programming Interface (API)

---

The REST API allows third-party applications to retrieve data stored within BIGSdb databases or to send new submissions to database curators. To use the REST API, your application will make a HTTP request and parse the response. The response format is JSON (except for routes that request a FASTA or CSV file).

Access to protected resources, i.e. those requiring an account, can be accessed via the API using *OAuth authentication*.

### 17.1 Passing additional/optional parameters

If you are using a method called with GET, optional parameters can be passed as arguments to the query URL by adding a '?' followed by the first argument and its value (separated by a '='). Additional parameters are separated by a '&', e.g.

[http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates?page=2&page\\_size=100](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates?page=2&page_size=100)

Methods called with POST require their arguments to be sent as JSON within the post body.

### 17.2 Paging using request headers

Paging of results can be selected using query parameters as described above. In this case, methods that support paging will include a paging object in the JSON response. This will contain links to the next page, last page etc.

The API also supports paging using request headers. The following request headers are supported:

- X-OFFSET
- X-PER-PAGE

e.g.

```
curl -i -H "X-PER-PAGE:10" -H "X-OFFSET:0" http://rest.pubmlst.org/db/pubmlst_
↪neisseria_isolates/isolates
```

If either of these headers are used, the paging object is no longer returned as part of the JSON response. The response will include the following headers:

- X-OFFSET
- X-PER-PAGE
- X-TOTAL-PAGES

## 17.3 Resources

- *GET / or /db* - List site resources
- *GET /db/{database}* - List database resources
- *GET /db/{database}/classification\_schemes* - List classification schemes
- *GET /db/{database}/classification\_schemes/{classification\_scheme\_id}* - Retrieve classification scheme information and groups
- *GET /db/{database}/classification\_schemes/{classification\_scheme\_id}/groups/{group\_id}* - List isolates belonging to a classification scheme group
- *GET /db/{database}/loci* - List loci
- *GET /db/{database}/loci/{locus}* - Retrieve locus record
- *GET /db/{database}/loci/{locus}/alleles* - Retrieve list of alleles defined for a locus
- *GET /db/{database}/loci/{locus}/alleles\_fasta* - Download alleles in FASTA format
- *GET /db/{database}/loci/{locus}/alleles/{allele\_id}* - Retrieve full allele information
- *POST /db/{database}/loci/{locus}/sequence* - Query sequence to identify allele
- *POST /db/{database}/sequence* - Query sequence to identify allele without specifying locus
- *GET /db/{database}/sequences* - Get summary of defined sequences
- *GET /db/{database}/schemes* - List schemes
- *GET /db/{database}/schemes/{scheme\_id}* - Retrieve scheme information
- *GET /db/{database}/schemes/{scheme\_id}/loci* - Retrieve scheme loci
- *GET /db/{database}/schemes/{scheme\_id}/fields/{field}* - Retrieve information about scheme field
- *GET /db/{database}/schemes/{scheme\_id}/profiles* - List allelic profiles defined for scheme
- *GET /db/{database}/schemes/{scheme\_id}/profiles\_csv* - Download allelic profiles in CSV (tab-delimited) format
- *GET /db/{database}/schemes/{scheme\_id}/profiles/{profile\_id}* - Retrieve allelic profile record
- *POST /db/{database}/schemes/{scheme\_id}/sequence* - Query sequence to extract allele designations/fields for a scheme
- *GET /db/{database}/isolates* - Retrieve list of isolate records
- *POST /db/{database}/isolates/search* - Search isolate database
- *GET /db/{database}/isolates/{isolate\_id}* - Retrieve isolate record
- *GET /db/{database}/isolates/{isolate\_id}/allele\_designations* - Retrieve list of allele designations
- *GET /db/{database}/isolates/{isolate\_id}/allele\_designations/{locus}* - Retrieve full allele designation record

- *GET /db/{database}/isolates/{isolate\_id}/allele\_ids* - Retrieve allele identifiers
- *GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele\_designations* - Retrieve scheme allele designation records
- *GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele\_ids* - Retrieve list of scheme allele identifiers
- *GET /db/{database}/isolates/{isolate\_id}/contigs* - Retrieve list of contigs
- *GET /db/{database}/isolates/{isolate\_id}/contigs\_fasta* - Download contigs in FASTA format
- *GET /db/{database}/isolates/{isolate\_id}/history* - Retrieve isolate update history
- *GET /db/{database}/contigs/{contig\_id}* - Retrieve contig record
- *GET /db/{database}/fields* - Retrieve list of isolate provenance field descriptions
- *GET /db/{database}/fields/{field}* - Retrieve values set for a provenance field
- *GET /db/{database}/users/{user\_id}* - Retrieve user information
- *GET /db/{database}/projects* - Retrieve list of projects
- *GET /db/{database}/projects/{project\_id}* - Retrieve project information
- *GET /db/{database}/projects/{project\_id}/isolates* - Retrieve list of isolates belonging to a project
- *GET /db/{database}/submissions* - Retrieve list of submissions
- *POST /db/{database}/submissions* - Create new submission
- *GET /db/{database}/submissions/{submission\_id}* - Retrieve submission record
- *DELETE /db/{database}/submissions/{submission\_id}* - Delete submission record
- *GET /db/{database}/submissions/{submission\_id}/messages* - Retrieve submission correspondence
- *POST /db/{database}/submissions/{submission\_id}/messages* - Add submission correspondence
- *GET /db/{database}/submissions/{submission\_id}/files* - retrieve list of supporting files uploaded for submission
- *POST /db/{database}/submissions/{submission\_id}/files* - Upload submission supporting file
- *GET /db/{database}/submissions/{submission\_id}/files/{filename}* - Download submission supporting file
- *DELETE /db/{database}/submissions/{submission\_id}/files/{filename}* - Delete submission supporting file

### 17.3.1 GET / or /db - List site resources

**Required route parameters:** None

**Optional query parameters:** None

**Example request URI:** <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

### 17.3.2 GET /db/{database} - List database resources

These will vary depending on whether the resource is an isolate or a sequence definition database.

**Required route parameter:** database [string] - Database configuration name

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates)

**Response:** Object containing a subset of the following key/value pairs:

- *fields* [string] - URI to isolate provenance field information
- *isolates* [string] - URI to isolate records
- *schemes* [string] - URI to list of schemes
- *loci* [string] - URI to list of loci
- *projects* [string] - URI to list of projects

### 17.3.3 GET /db/{database}/classification\_schemes - List classification schemes

**Required route parameter:** database [string] - Database configuration name

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/classification\\_schemes](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/classification_schemes)

**Response:** Object containing:

- records [int] - Number of classification schemes.
- classification\_schemes [array] - List of *URIs to classification schemes*.

### 17.3.4 GET /db/{database}/classification\_schemes/{classification\_scheme\_id} - Retrieve classification scheme information and groups

Sequence definition databases only.

**Required route parameters:**

- database [string] - Database configuration name
- classification\_scheme\_id [integer] - Classification scheme id number

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/classification\\_schemes/1](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/classification_schemes/1)

**Response:** Object containing some or all of:

- id [integer] - Classification scheme id
- name [text] - Name of classification scheme
- description [text] - Description of classification scheme
- relative\_threshold [boolean] - True if a *relative threshold* is used
- inclusion\_threshold [integer] - The threshold for number of loci difference used to group

- groups [array] (sequence definition databases only) - list of group objects consisting of:
  - id [integer] - group id
  - profiles [array] - list of *URLs to profiles* belonging to the group

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

Isolate databases only.

#### Required route parameters:

- database [string] - Database configuration name
- classification\_scheme\_id [integer] - Classification scheme id number
- group\_id [integer] - Group id number

#### Optional parameters:

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/classification\\_schemes/4/groups/65](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/classification_schemes/4/groups/65)

**Response:** Object containing of:

- records [integer] - Number of isolates
- isolates [array] - List of *URLs to isolate records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.6 GET /db/{database}/loci - List loci

**Required route parameter:** database [string] - Database configuration name

#### Optional parameters:

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.
- alleles\_added\_after [date] - Include only loci with alleles added after specified date (ISO 8601 format). Only recognized in sequence definition databases.
- alleles\_updated\_after [date] - Include only loci with alleles last modified after specified date (ISO 8601 format). Only recognized in sequence definition databases.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/loci](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci)

**Response:** Object containing:

- records [integer] - Number of loci
- loci [array] - List of *URIs to defined locus records*. Pages are 100 records by default. Page size can be modified using the `page_size` parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

---

**Note:** See also the *scheme specific version*, allowing filtering by date of last allele update for just the loci that are members of a scheme.

---

### 17.3.7 GET /db/{database}/loci/{locus} - Retrieve locus record

Provides information about a locus, including links to allele sequences (in seqdef databases).

**Required route parameters:**

- database [string] - Database configuration name
- locus [string] - Locus name

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/loci/abcZ](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)

### 17.3.8 GET /db/{database}/loci/{locus}/alleles - Retrieve list of alleles defined for a locus

#### Required route parameters:

- database [string] - Database configuration name
- locus [string] - Locus name

#### Optional parameters:

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.
- added\_after [date] - Include only alleles added after specified date (ISO 8601 format).
- added\_on [date] - Include only alleles added on specified date (ISO 8601 format).
- updated\_after [date] - Include only alleles last modified after specified date (ISO 8601 format).
- updated\_on [date] - Include only alleles last modified on specified date (ISO 8601 format).

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/loci/abcZ/alleles](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci/abcZ/alleles)

**Response:** Object containing:

- records [int] - Number of alleles.

- last\_updated [date] - Latest allele addition/modification date (ISO 8601 format).
- alleles [array] - List of *URIs to defined allele records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.9 GET /db/{database}/loci/{locus}/alleles\_fasta - Download alleles in FASTA format

#### Required route parameters:

- database [string] - Database configuration name
- locus [string] - Locus name

#### Optional parameters:

- added\_after [date] - Include only alleles added after specified date (ISO 8601 format).
- added\_on [date] - Include only alleles added on specified date (ISO 8601 format).
- updated\_after [date] - Include only alleles last modified after specified date (ISO 8601 format).
- updated\_on [date] - Include only alleles last modified on specified date (ISO 8601 format).

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/loci/abcZ/alleles\\_fasta](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/loci/abcZ/alleles_fasta)

**Response:** FASTA format file of allele sequences

### 17.3.10 GET /db/{database}/loci/{locus}/alleles/{allele\_id} - Retrieve full allele information

#### Required route parameters:

- database [string] - Database configuration name
- locus [string] - Locus name
- allele\_id [string] - Allele identifier

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/loci/abcZ/alleles/5](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)
- timestamp [string] - last updated date (ISO 8601 format)

### 17.3.11 POST /db/{database}/loci/{locus}/sequence - Query sequence to identify allele

#### Required route parameters:

- database [string] - Database configuration name
- locus [string] - Locus name

#### Required additional parameters (JSON-encoded in POST body):

- sequence [string] - Sequence string or base64-encoded FASTA file

#### Optional parameters (JSON-encoded in POST body):

- details [true/false] - Return detailed exact match parameters
- base64 [true/false] - Sequence is a base64-encoded FASTA file

#### Response: Object containing the following key/value pairs:

- exact\_matches [array] - list of match objects, each consisting of:
  - allele\_id
  - href - *URI to allele record*.

additionally if 'details' parameter passed:

- start - start position on query
- end - end position on query
- orientation - forward/reverse
- length - length of matched allele
- contig - contig name if FASTA file is uploaded

If the locus is linked to field data in client isolate databases, there may also be an object called 'linked\_data' containing values and frequencies of the field for the returned allele.

- best\_match [object] - consisting of key/value pairs (if no exact matches)
  - allele\_id
  - href - *URI to allele record*.
  - start - start position on query (predicted taking account of allele length)
  - end - end position on query (predicted taking account of allele length)
  - orientation - forward/reverse
  - length - length of matched allele
  - alignment - length of BLAST alignment
  - mismatches - number of mismatches

- identity - %identity of match
- gaps - number of gaps in alignment

### 17.3.12 POST /db/{database}/sequence - Query sequence to identify allele without specifying locus

**Required route parameters:**

- database [string] - Database configuration name

**Required additional parameters (JSON-encoded in POST body):**

- sequence [string] - Sequence string or base64-encoded FASTA file

**Optional parameters (JSON-encoded in POST body):**

- details [true/false] - Return detailed exact match parameters
- base64 [true/false] - Sequence is a base64-encoded FASTA file

**Response:**

- exact\_matches [object] consisting of locus keys, each consisting of array of match objects consisting of:
  - allele\_id
  - href - *URI to allele record*.

additionally if 'details' parameter passed:

- start - start position on query
- end - end position on query
- orientation - forward/reverse
- length - length of matched allele
- contig - contig name if FASTA file is uploaded

If the locus is linked to field data in client isolate databases, there may also be an object called 'linked\_data' containing values and frequencies of the field for the returned allele.

---

**Note:** This method only supports exact matches. If no match is indicated for a specific locus, use the *locus-specific call* to identify the closest match.

---

### 17.3.13 GET /db/{database}/sequences - Get summary of defined sequences

**Required route parameter:** database [string] - Database configuration name

**Optional parameters:**

- added\_after [date] - Count only alleles added after specified date (ISO 8601 format).
- added\_on [date] - Count only alleles added on specified date (ISO 8601 format).
- updated\_after [date] - Count only alleles last modified after specified date (ISO 8601 format).
- updated\_on [date] - Count only allele updated on specified date (ISO 8601 format).

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/sequences](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/sequences)

**Response:** Object containing a subset of the following key/value pairs:

- *loci* [string] - URI to list of loci
- records [int] - Number of alleles defined
- last\_updated [date] - Latest allele addition/modification date (ISO 8601 format).

### 17.3.14 GET /db/{database}/schemes - List schemes

**Required route parameter:** database [string] - Database configuration name

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/schemes](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes)

**Response:**

- records [integer] - Number of schemes
- schemes [array] - list of scheme objects, each containing:
  - scheme [string] - *URI to scheme information*
  - description [string]

### 17.3.15 GET /db/{database}/schemes/{scheme\_id} - Retrieve scheme information

Includes links to allelic profiles (in seqdef databases, if appropriate). **Required route parameters:**

- database [string] - Database configuration name
- scheme\_id [integer] - Scheme id number

**Optional parameters:**

- added\_after [date] - Count only profiles added after specified date (ISO 8601 format).
- added\_on [date] - Count only profiles added on specified date (ISO 8601 format).
- updated\_after [date] - Count only profiles last modified after specified date (ISO 8601 format).
- updated\_on [date] - Count only profiles updated on specified date (ISO 8601 format).

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/schemes/1](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1)

**Response:** Object containing a subset of the following key/value pairs:

- id [integer]
- description [string]
- locus\_count [integer] - number of loci belonging to scheme
- loci [array] - list of *URIs to locus descriptions*
- has\_primary\_key\_field [boolean]
- fields [array] - list of *URIs to scheme field descriptions*
- primary\_key\_field [string] - *URI to primary key field description*
- profiles [string] - URI to list of profile definitions (only seqdef databases)
- profiles\_csv [string] - URI to tab-delimited file of all scheme profiles

- curators [array] (seqdef databases) - list of *URIs to user records* of curators of the scheme
- records [int] - Number of profiles
- last\_added [date] - Latest profile addition/modification date (ISO 8601 format).
- last\_updated [date] - Latest profile addition/modification date (ISO 8601 format).

### 17.3.16 GET /db/{database}/schemes/{scheme\_id}/loci - Retrieve scheme loci

**Required route parameters:**

- database [string] - Database configuration name
- scheme\_id [integer] - Scheme id number

**Optional parameters:**

- alleles\_added\_after [date] - Include only loci with alleles added after specified date (ISO 8601 format). Only recognized in sequence definition databases.
- alleles\_updated\_after [date] - Include only loci with alleles last modified after specified date (ISO 8601 format). Only recognized in sequence definition databases.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/schemes/1/loci](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/loci)

**Response:** Object containing:

- records [integer] - Number of loci
- loci [array] - List of *URIs to defined locus records*.

### 17.3.17 GET /db/{database}/schemes/{scheme\_id}/fields/{field} - Retrieve information about scheme field

**Required route parameters:**

- database [string] - Database configuration name
- scheme\_id [integer] - Scheme id number
- field [string] - Field name

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/schemes/1/fields/ST](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

### 17.3.18 GET /db/{database}/schemes/{scheme\_id}/profiles - List allelic profiles defined for scheme

**Required route parameters:**

- database [string] - Database configuration name

- `scheme_id` [integer] - Scheme id

**Optional parameters:**

- `page` [integer]
- `page_size` [integer]
- `return_all` [integer] - Set to non-zero value to disable paging.
- `added_after` [date] - Include only profiles added after specified date (ISO 8601 format).
- `added_on` [date] - Include only profiles added on specified date (ISO 8601 format).
- `updated_after` [date] - Include only profiles last modified after specified date (ISO 8601 format).
- `updated_on` [date] - Include only profiles last modified on specified date (ISO 8601 format).

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/schemes/1/profiles](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/profiles)

**Response:** Object containing:

- `records` [int] - Number of profiles
- `last_updated` [date] - Latest profile addition/modification date (ISO 8601 format).
- `profiles` [array] - List of *URIs to defined profile records*. Pages are 100 records by default. Page size can be modified using the `page_size` parameter.
- `paging` [object] - Some or all of the following:
  - `previous` - URI to previous page of results
  - `next` - URI to next page of results
  - `first` - URI to first page of results
  - `last` - URI to last page of results
  - `return_all` - URI to page containing all results (paging disabled)

---

**Note:** This method also supports content negotiation. If the request accepts header includes TSV or CSV, then the call is redirected to `/db/{database}/schemes/{scheme_id}/profiles_csv`.

---

### 17.3.19 GET `/db/{database}/schemes/{scheme_id}/profiles_csv` - Download allelic profiles in CSV (tab-delimited) format

**Required route parameters:**

- `database` [string] - Database configuration name
- `scheme_id` [integer] - Scheme id

**Optional parameters:**

- `added_after` [date] - Include only profiles added after specified date (ISO 8601 format).
- `added_on` [date] - Include only profiles added on specified date (ISO 8601 format).
- `updated_after` [date] - Include only profiles last modified after specified date (ISO 8601 format).
- `updated_on` [date] - Include only profiles last modified on specified date (ISO 8601 format).

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/schemes/1/profiles\\_csv](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/schemes/1/profiles_csv)

**Response:** Tab-delimited text file of allelic profiles

### 17.3.20 GET /db/{database}/schemes/{scheme\_id}/profiles/{profile\_id} - Retrieve allelic profile record

**Required route parameters:**

- database [string] - Database configuration name
- scheme\_id [integer] - Scheme id
- profile\_id [string/integer] - Profile id

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/schemes/1/profiles/11](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)

### 17.3.21 POST /db/{database}/schemes/{scheme\_id}/sequence - Query sequence to extract allele designations/fields for a scheme

**Required route parameters:**

- database [string] - Database configuration name
- scheme\_id [integer] - Scheme id

**Required additional parameters (JSON-encoded in POST body):**

- sequence [string] - Sequence string or base64-encoded FASTA file

**Optional parameters (JSON-encoded in POST body):**

- details [true/false] - Return detailed exact match parameters
- base64 [true/false] - Sequence is a base64-encoded FASTA file

**Response:** Object containing the following key/value pairs:

- exact\_matches [array] - list of match objects, each consisting of:
  - allele\_id
  - href - *URI to allele record*.

additionally if 'details' parameter passed:



- start - start position on query
- end - end position on query
- orientation - forward/reverse
- length - length of matched allele
- contig - contig name if FASTA file is uploaded

If the locus is linked to field data in client isolate databases, there may also be an object called 'linked\_data' containing values and frequencies of the field for the returned allele.

Example curl call to upload a FASTA file 'contigs.fasta' and extract MLST results from Neisseria database:

```
(echo -n '{"base64":true,"sequence": "' ; base64 contigs.fasta; echo '"}') |
curl -s -H "Content-Type: application/json" -X POST "http://rest.pubmlst.org/db/
↪pubmlst_neisseria_seqdef/schemes/1/sequence" -d @-
```

**Note:** This method only supports exact matches. If no match is indicated for a specific locus, use the *locus-specific call* to identify the closest match.

### 17.3.22 GET /db/{database}/isolates - Retrieve list of isolate records

**Required route parameter:** database [string] - Database configuration name

**Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.
- added\_after [date] - Include only isolates added after specified date (ISO 8601 format).
- added\_on [date] - Include only isolates added on specified date (ISO 8601 format).
- updated\_after [date] - Include only isolates last modified after specified date (ISO 8601 format).
- updated\_on [date] - Include only isolates updated on specified date (ISO 8601 format).

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates)

**Response:** Object containing:

- records [int] - Number of isolates
- isolates [array] - List of *URIs to isolate records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.23 GET /db/{database}/isolates/{isolate\_id} - Retrieve isolate record

**Required route parameters:**

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier

**Optional parameter:**

- provenance\_only [integer] - Set to non-zero value to only return provenance metadata

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1](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 some of the following:
  - description [string] - description of scheme
  - loci\_designated\_count [integer] - number of loci within scheme that have an allele designated for this isolate.
  - allele\_ids [string] - *URI to list of all allele\_id values defined for this scheme* for this isolate
  - full\_designations [string] - *URI to list of full allele designation records* for this isolate
  - fields [object] - consisting of key/value pairs where the key is the name of each scheme field
  - classification\_schemes [object] - consisting of key/value pairs, where each key is the name of the classification scheme and the value is an object consisting of:
    - \* href [string] - *URI to classification scheme description*
    - \* groups [array] - list of group objects consisting of:
      - group [integer] - group id

- records [integer] - number of isolates in group
- isolates [string] - URI to classification group record containing URIs to member isolate records
- projects [array] - list of project objects, each containing the following:
  - id [string] - *URI to project information*
  - description [string] - description of project
- history [string] - *URI to isolate history record*
- new\_version [string] - URI to newer version of record
- old\_version [string] - URI to older version of record

### 17.3.24 GET /db/{database}/isolates/{isolate\_id}/allele\_designations - Retrieve list of allele designation records

#### Required route parameters:

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier

#### Optional parameters:

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1/allele\\_designations](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/allele_designations)

**Response:** Object containing:

- records [int] - Number of allele designations
- allele\_designations [array] - List of *URIs to allele designation records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.25 GET /db/{database}/isolates/{isolate\_id}/allele\_designations/{locus} - Retrieve full allele designation record

#### Required route parameters:

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier
- locus [string] - Locus name

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1/allele\\_designations/BACT000065](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)

### 17.3.26 GET /db/{database}/isolates/{isolate\_id}/allele\_ids - Retrieve allele identifiers

**Required route parameters:**

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier

**Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1/allele\\_ids](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/allele_ids)

**Response:** Object containing:

- records [int] - Number of allele id objects
- allele\_ids [array] - List of allele id objects, each consisting of a key/value pair where the key is the locus name. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.27 GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele\_designations - Retrieve scheme allele designation records

**Required route parameters:**

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier
- scheme\_id [integer] - Scheme identifier

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1/schemes/1/allele\\_designations](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/schemes/1/allele_designations)

**Response:**

- records [int] - Number of allele designation objects
- allele\_designations [array] - List of *allele designation objects* for each locus in the specified scheme that has been designated.

### 17.3.28 GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele\_ids - Retrieve list of scheme allele identifiers

**Required route parameters:**

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier
- scheme\_id [integer] - Scheme identifier

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1/schemes/1/allele\\_ids](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/schemes/1/allele_ids)

**Response:**

- records [int] - Number of allele id objects
- allele\_ids [array] - List containing allele id objects for each locus in the specified scheme that has been designated. Each allele\_id object contains a key which is the name of the locus with a value that may be either a string, integer or array of strings or integers (required where there are multiple designations for a locus). The data type depends on the allele\_id\_format set for the specific locus.

### 17.3.29 GET /db/{database}/isolates/{isolate\_id}/contigs - Retrieve list of contigs

**Required route parameters:**

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier

**Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1/contigs](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/contigs)

**Response:** Object containing:

- records [int] - Number of contigs
- contigs [array] - List of *URLs to contig records* Pages are 100 records by default. Page size can be modified using the `page_size` parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.30 GET /db/{database}/isolates/{isolate\_id}/contigs\_fasta - Download contigs in FASTA format

**Required route parameters:**

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier

**Optional parameter:**

- header [string] - either 'original\_designation' or 'id' (default is 'id'). This selects whether the FASTA header lines contain the originally uploaded FASTA headers or the sequence bin id numbers.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1/contigs\\_fasta?header=original\\_designation](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/contigs_fasta?header=original_designation)

**Response:** FASTA format file of isolate contig sequences

### 17.3.31 GET /db/{database}/isolates/{isolate\_id}/history - Retrieve isolate update history

**Required route parameters:**

- database [string] - Database configuration name
- isolate\_id [integer] - Isolate identifier

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/isolates/1/history](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/1/history)

**Response:** Object containing:

- records [int] - Number of updates
- contigs [array] - List of update objects each consisting of the following key/value pairs:
  - curator [string] - *URI to user details* of curator
  - timestamp [string] - Time of update
  - actions [array] - List of update descriptions [strings]

### 17.3.32 POST /db/{database}/isolates/search - Search isolate database

**Required route parameters:**

- database [string] - Database configuration name

**Optional parameters (appended to URI):**

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.

**Query parameters (JSON-encoded in POST body):**

You must include at least one query parameter.

Flattened parameter names in the following forms are supported:

- field.{field} - key/value pairs for provenance fields. Supported field names can be found by calling the */fields route*. The fields will vary by database.
- locus.{locus} - key/value pairs of locus and its allele designation. Supported locus names can be found by calling the */loci route*.
- scheme.{scheme\_id}.{scheme\_field} - key/value pairs of scheme fields and their values. Supported field names can be determined by following routes from the */schemes route*.

Parameters have to be flattened rather than using nested attributes to support passing via OAuth.

**Example method call using curl:** The following searches for *Neisseria* ST-11 isolates from Europe in 2015 (MLST is scheme#1 in this database).

```
curl -s -H "Content-Type: application/json" -X POST \
"http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/isolates/search" \
-d '{"field.continent":"europe","field.year":2015,"scheme.1.ST":11}'
```

**Response:** Object containing:

- records [int] - Number of isolates
- isolates [array] - List of *URIs to isolate records*. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.33 GET /db/{database}/contigs/{contig\_id} - Retrieve contig record

**Required route parameters:**

- database [string] - Database configuration name
- contig\_id [integer] - Contig identifier

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/contigs/180062](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)
- loci [array] - list of sequence tag objects consisting of:
  - locus [string] - *URI to locus description*
  - locus\_name [string]
  - start [integer]
  - end [integer]
  - direction [string] - forward/reverse
  - complete [boolean] - true/false

### 17.3.34 GET /db/{database}/fields - Retrieve list of isolate provenance field descriptions

**Required route parameters:**

- database [string] - Database configuration name

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/fields](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



- values [string] - URI to list of used field values

### 17.3.35 GET /db/{database}/fields/{field} - Retrieve values set for a provenance field

**Required route parameters:**

- database [string] - Database configuration name
- field [string] - Provenance metadata field name

**Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/fields/country](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/fields/country)

**Response:** Object containing:

- records [int] - Number of values
- values [array] - List of values used in isolate records. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.36 GET /db/{database}/users/{user\_id} - Retrieve user information

Users may be data submitters or curators.

**Required route parameters:**

- database [string] - Database configuration name
- user\_id [integer] - User id number

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/users/2](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

### 17.3.37 GET /db/{database}/projects - Retrieve list of projects

**Required route parameter:** database [string] - Database configuration name

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/projects](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/projects)

**Response:**

- projects [array] - List of project objects, each containing:
  - project [string] - *URI to project information*
  - description [string]
  - isolate\_count [integer] - number of isolates in project

### 17.3.38 GET /db/{database}/projects/{project\_id} - Retrieve project information

**Required route parameters:**

- database [string] - Database configuration name
- project\_id [integer] - Project id number

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/projects/3](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/projects/3)

**Response:** Object containing a subset of the following key/value pairs:

- id [integer]
- description [string]
- isolates [string] - *URI to list of URIs of member isolate records.*

### 17.3.39 GET /db/{database}/projects/{project\_id}/isolates - Retrieve list of isolates belonging to a project

**Required route parameter:**

- database [string] - Database configuration name
- project\_id [integer] - Project id number

**Optional parameters:**

- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/projects/3/isolates](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/projects/3/isolates)

**Response:** Object containing:

- records [int] - Number of isolates in the project
- isolates [array] - List of URIs to isolate records. Pages are 100 records by default. Page size can be modified using the page\_size parameter.
- paging [object] - Some or all of the following:

- previous - URI to previous page of results
- next - URI to next page of results
- first - URI to first page of results
- last - URI to last page of results
- return\_all - URI to page containing all results (paging disabled)

### 17.3.40 GET /db/{database}/submissions - retrieve list of submissions

**Required route parameter:** database [string] - Database configuration name

**Optional parameters:**

- type [string] - either 'alleles', 'profiles' or 'isolates'
- status [string] - either 'closed' or 'pending'
- page [integer]
- page\_size [integer]
- return\_all [integer] - Set to non-zero value to disable paging.

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_isolates/submissions](http://rest.pubmlst.org/db/pubmlst_neisseria_isolates/submissions)

**Response:** Object containing:

- records [int] - Number of submissions
- submissions [array] - List of *URIs to submission records*
- paging [object] - Some or all of the following:
  - previous - URI to previous page of results
  - next - URI to next page of results
  - first - URI to first page of results
  - last - URI to last page of results
  - return\_all - URI to page containing all results (paging disabled)

### 17.3.41 POST /db/{database}/submissions - create new submission

**Required route parameter:** database [string] - Database configuration name

**Required additional parameters (JSON-encoded in POST body):**

- type [string] - either:
  - alleles (sequence definition databases only)
  - profiles (sequence definition databases only)
  - isolates (isolate databases only)
  - genomes (isolate databases only)

The following are required with the specified database type:

#### Allele submissions

- locus [string] - name of locus

- technology [string] - name of sequencing technology: either '454', 'Illumina', 'Ion Torrent', 'PacBio', 'Oxford Nanopore', 'Sanger', 'Solexa', 'SOLiD', or 'other'
- read\_length [string] - read length of sequencing: either '<100', '100-199', '200-299', '300-499', '>500', or any positive integer (only required for Illumina)
- coverage [string] - mean coverage of sequencing: either '<20x', '20-49x', '50-99x', '>100x', or any positive integer (only required for Illumina)
- assembly [string] - assembly method: either 'de novo' or 'mapped'
- software [string] - name of assembly software
- sequences [string] - either single raw sequence or multiple sequences in FASTA format

#### Profile submissions

- scheme\_id [int] - scheme id number
- profiles [string] - tab-delimited profile data - this should include a header line containing the name of each locus

#### Isolate submissions

- isolates [string] - tab-delimited isolate data - this should include a header line containing each field or locus included

#### Genome submissions

- isolates [string] - tab-delimited isolate data - this should include a header line containing each field or locus included as well as for 'assembly\_filename' and 'sequence\_method'. The 'sequence\_method' should be either '454', 'Illumina', 'Ion Torrent', 'PacBio', 'Oxford Nanopore', 'Sanger', 'Solexa', 'SOLiD', or 'other'. Following submission, contig files should be uploaded with the same names as set for 'assembly\_filename'. This can be done using the [file upload route](#).

#### Optional parameters:

- message [string] - correspondence to the curator
- email [int] - set to 1 to enable E-mail updates (E-mails will be sent to the registered user account address).

#### Response: Object containing:

- submission - [URI to submission record](#)

For genome submissions, the response object will also contain:

- missing\_files [array] - List of filenames that need to be uploaded to complete the submission. These filenames are defined in the 'assembly\_filename' field of the isolate record upload. The files should contain the contig assemblies.
- message [string] - 'Please upload missing contig files to complete submission.'

### 17.3.42 GET /db/{database}/submissions/{submission\_id} - Retrieve submission record

#### Required route parameters:

- database [string] - Database configuration name
- submission\_id [string] - Submission id

#### Optional parameters: None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/submissions/BIGSdb\\_20151013081836\\_14559\\_14740](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/submissions/BIGSdb_20151013081836_14559_14740)

**Response:** Object containing some of the following:

- id [string] - Submission id
- type [string] - Either 'alleles', 'profiles', 'isolates'
- date\_submitted [string] - Submission date (ISO 8601 format)
- timestamp [string] - Last updated date (ISO 8601 format)
- submitter [string] - *URI to user details* of submitter
- curator [string] - *URI to user details* of curator
- status [string] - either 'started', 'pending', or 'closed'
- outcome [string] - either 'good' (data uploaded), 'bad' (data rejected), or 'mixed' (parts of submission accepted)
- correspondence [array] - List of correspondence objects in time order. Each contains:
  - user [string] *URI to user details* of user
  - timestamp [string]
  - message [string]

#### Allele submissions

- locus [string] - name of locus
- technology [string] - name of sequencing technology: either '454', 'Illumina', 'Ion Torrent', 'PacBio', 'Oxford Nanopore', 'Sanger', 'Solexa', 'SOLiD', or 'other'
- read\_length [string] - read length of sequencing: either '<100', '100-199', '200-299', '300-499', '>500', or any positive integer (only required for Illumina)
- coverage [string] - mean coverage of sequencing: either '<20x', '20-49x', '50-99x', '>100x', or any positive integer (only required for Illumina)
- assembly [string] - assembly method: either 'de novo' or 'mapped'
- software [string] - name of assembly software
- seqs [array] - List of sequence objects each containing:
  - seq\_id [string] - Sequence identifier
  - assigned\_id [string] - Allele identifier if uploaded to the database (otherwise undefined)
  - status [string] - Either 'pending', 'assigned', or 'rejected'
  - sequence [string]

#### Profile submissions

- scheme [string] - *URI to scheme information*
- profiles [array] - List of profile record objects. Each contains:
  - profile\_id [string] - Record identifier
  - assigned\_id [string] - Profile identifier if uploaded to the database (otherwise undefined)
  - status [string] - Either 'pending', 'assigned', or 'rejected'
  - designations [object] containing key/value pairs for each locus containing the allele identifier

#### Isolate submissions

- isolates [array] - List of isolate record objects. Each contains key/value pairs for included fields.

### 17.3.43 DELETE /db/{database}/submissions/{submission\_id} - Delete submission record

You must be the owner and the record must be closed.

**Required route parameters:**

- database [string] - Database configuration name
- submission\_id [string] - Submission id

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/submissions/BIGSdb\\_20151013081836\\_14559\\_14740](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/submissions/BIGSdb_20151013081836_14559_14740)

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

### 17.3.44 GET /db/{database}/submissions/{submission\_id}/messages - Retrieve submission correspondence

**Required route parameters:**

- database [string] - Database configuration name
- submission\_id [string] - Submission id

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/submissions/BIGSdb\\_20151013081836\\_14559\\_14740/messages](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/submissions/BIGSdb_20151013081836_14559_14740/messages)

**Response:** Array of correspondence objects in time order. Each contains:

- user [string] *URI to user details* of user
- timestamp [string]
- message [string]

### 17.3.45 POST /db/{database}/submissions/{submission\_id}/messages - Add submission correspondence

**Required route parameters:**

- database [string] - Database configuration name
- submission\_id [string] - Submission id

**Required additional parameter (JSON-encoded in POST body):**

- message [string] - Message text

**Optional parameters:** None

**Response:** message [string] - 'Message added.'

### 17.3.46 GET /db/{database}/submissions/{submission\_id}/files - Retrieve list of supporting files uploaded for submission

**Required route parameters:**

- database [string] - Database configuration name
- submission\_id [string] - Submission id

**Optional parameters:** None

**Example request URI:** [http://rest.pubmlst.org/db/pubmlst\\_neisseria\\_seqdef/submissions/BIGSdb\\_20151013081836\\_14559\\_14740/files](http://rest.pubmlst.org/db/pubmlst_neisseria_seqdef/submissions/BIGSdb_20151013081836_14559_14740/files)

**Response:** Array of URIs to files

### 17.3.47 POST /db/{database}/submissions/{submission\_id}/files - Upload submission supporting file

**Required route parameters:**

- database [string] - Database configuration name
- submission\_id [string] - Submission id

**Required additional parameters (JSON-encoded in POST body):**

- filename [string] - Name of file to store within submission
- upload [base64 encoded data] - Raw file data

**Optional parameters:** None

**Response:** message [string] - 'File uploaded.'

### 17.3.48 GET /db/{database}/submissions/{submission\_id}/files/{filename} - Download submission supporting file

**Required route parameters:**

- database [string] - Database configuration name
- submission\_id [string] - Submission id
- filename [string] - Name of file

**Optional parameters:** None

**Response:** File download

### 17.3.49 DELETE /db/{database}/submissions/{submission\_id}/files/{filename} - Delete submission supporting file

**Required route parameters:**

- database [string] - Database configuration name
- submission\_id [string] - Submission id
- filename [string] - Name of file

**Optional parameters:** None

**Response:** message [string] - 'File deleted.'

## 17.4 Authentication

Protected resources, i.e. those requiring a user to log in, can be accessed via the API using OAuth (1.0A) authentication (see [IETF RFC5849](#) for details). Third-party client software has to be registered with the BIGSdb site before they can access authenticated resources. The overall three-legged flow works as follows:

1. *Developer signs up* and gets a consumer key and consumer secret specific to their application.
2. Application *gets a request token* and directs user to authorization page on BIGSdb.
3. BIGSdb *asks user for authorization* for application to access specific resource using their credentials. A verifier code is provided.
4. The application exchanges the request token and OAuth verifier code for an *access token and secret* (these do not expire but may be revoked by the user or site admin).
5. Application uses access token/secret to *request session token* (this is valid for 12 hours).
6. All calls to *access protected resources* are signed using the session token/secret and consumer key/secret.

It is recommended that application developers use an OAuth library to generate and sign requests.

### 17.4.1 Developer sign up to get a consumer key

Application developers should apply to the site administrator of the site running BIGSdb. The administrator can *generate a key and secret* using a script - both of these will need to be used by the application to sign requests.

The client id is usually a 24 character alphanumeric string. The secret is usually a 42 character alphanumeric (including punctuation) string, e.g.

- **client\_id:** efKXmqp2D0EB1MBkZaGC2lPf
- **client\_secret:** F\$M+fQ2AFFB2YBDfF9fpHF^qSWJdmmN%L4Fxf5Gur3

### 17.4.2 Getting a request token

- **Relative URL:** /db/{database}/oauth/get\_request\_token
- **Supported method:** GET

The application uses the consumer key to obtain a request token. The request token is a temporary token used to initiate user authorization for the application and will expire in 60 minutes. The request needs to contain the following parameters and to be signed using the consumer secret:

- oauth\_consumer\_key
- oauth\_request\_method ('GET')
- oauth\_request\_url (request URL)
- oauth\_signature\_method ('HMAC-SHA1')
- oauth\_signature
- oauth\_timestamp (UNIX timestamp - seconds since Jan 1 1970) - this must be within 600 seconds of the current time.



- `oauth_callback` ('oob' for desktop applications)
- `oauth_nonce` (random string)
- `oauth_version` ('1.0')

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

- **`oauth_token`**
  - This is the request token. It is usually a 32 character alphanumeric string.
  - e.g. `fKFm0WNhCfbEX8zQm6qhDA8K23FOWDGE`
- **`oauth_token_secret`**
  - This is the secret associated with the request token. It is usually a 32 character alphanumeric string.
  - e.g. `aZ0fncP7i5w5jlebdK5zyQ4vrRRVcdnv`
- **`oauth_callback_confirmed`**
  - This parameter is always set to true.

### 17.4.3 Getting user authorization

Once a request token has been obtained, this can be used by the end user to grant permission to access a specific resource to the application. The application should direct the user to the client authorization page (`authorizeClient`) specific to a database within BIGSdb, e.g. [http://pubmlst.org/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst\\_neisseria\\_seqdef&page=authorizeClient&oauth\\_token=fKFm0WNhCfbEX8zQm6qhDA8K23FOWDGE](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.

### 17.4.4 Getting an access token

- **Relative URL:** /db/{database}/oauth/get\_access\_token
- **Supported method:** GET

The application uses the request token, verifier code and its consumer key to obtain an access token. The access token does not expire but can be revoked by both the end user or the site administrator. The request needs to contain the following parameters and to be signed using the consumer secret and request token secret:

- `oauth_consumer_key`
- `oauth_request_method` ('GET')
- `oauth_request_url` (request URL)
- `oauth_signature_method` ('HMAC-SHA1')
- `oauth_signature`
- `oauth_token` (request token)
- `oauth_timestamp` (UNIX timestamp - seconds since Jan 1 1970) - this must be within 600 seconds of the current time.
- `oauth_nonce` (random string)
- `oauth_version` ('1.0')

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

- **oauth\_token**
  - This is the access token. It is usually a 32 character alphanumeric string.
  - e.g. SDrC74ZVl5SYSqY8lWZqrRxnyDnNGVFO
- **oauth\_token\_secret**
  - This is the secret associated with the access token. It is usually a 32 character alphanumeric string.
  - e.g. tYl2SPzgiO02lRVzW4JR1ez6Vvm4gVyv

### 17.4.5 Getting a session token

- **Relative URL:** /db/{database}/oauth/get\_session\_token
- **Supported method:** GET

The application uses the access token and its consumer key to obtain a session token. The session token is valid for 12 hours before it expires. The request needs to contain the following parameters and to be signed using the consumer secret and access token secret:

- **oauth\_consumer\_key**
- **oauth\_request\_method** ('GET')
- **oauth\_request\_url** (request URL)
- **oauth\_signature\_method** ('HMAC-SHA1')
- **oauth\_signature**
- **oauth\_token** (access token)
- **oauth\_timestamp** (UNIX timestamp - seconds since Jan 1 1970) - this must be within 600 seconds of the current time.
- **oauth\_nonce** (random string)
- **oauth\_version** ('1.0')

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

- **oauth\_token**
  - This is the session token. It is usually a 32 character alphanumeric string.
  - e.g. H8CjIS8Ikq6hwCUqUfF1l4pTaCYl8Ljw
- **oauth\_token\_secret**
  - This is the secret associated with the session token. It is usually a 32 character alphanumeric string.
  - e.g. RfponbaNPO7tkZ2miHFISk0pMndePNfJ

### 17.4.6 Accessing protected resources

- **Supported method:** GET

The application uses the session token and its consumer key to access a protected resource. The request needs to contain the following parameters and to be signed using the consumer secret and session token secret:

- **oauth\_consumer\_key**

- `oauth_request_method` ('GET')
- `oauth_request_url` (request URL)
- `oauth_signature_method` ('HMAC-SHA1')
- `oauth_signature`
- `oauth_token` (session token)
- `oauth_timestamp` (UNIX timestamp - seconds since Jan 1 1970) - this must be within 600 seconds of the current time.
- `oauth_nonce` (random string)
- `oauth_version` ('1.0')

---

Frequently asked questions (FAQs)

---

## 18.1 General

### 1. What is the minimum specification of hardware required to run BIGSdb?

The software will run on fairly 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 multiple cores (depending on the settings in `bigsdbs.conf`), so if you want to run multiple jobs in parallel then you may want more cores (and memory). Tagging of new genomes using the offline *autotagger* is usually run in multi-threaded mode so the more cores available the faster this will be.

As a comparison, the PubMLST site is run on two machines - separate web and database servers. All offline jobs and tagging of genomes is performed on the database server. These have the following specification:

- web server: 40 cores, 128GB RAM
- database server: 64 cores, 1TB RAM, 3TB RAID 10 local storage

### 2. Why might icons be missing when using Internet Explorer?

This can occur if you have Compatibility Mode enabled. BIGSdb generates valid HTML5 and Compatibility Mode should not be used. Please ensure this is not enabled in the Internet Explorer tools section.

## 18.2 Installation

### 1. BIGSdb is accumulating files in various temp directories - is this normal and how do I clean them out?

See: *Periodically delete temporary files.*

### 2. BIGSdb is complaining of an invalid script path - what does this mean?

In your database config.xml file system tag are two attributes - script\_path\_includes and curate\_path\_includes. These contain regexes that the web url to your script (bigsdbs.pl and bigscurate.pl respectively) must match. This prevents somebody from accessing a private database using an instance of bigsdbs.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/bigsdbs/bigsdbs.pl> then you can set script\_path\_includes to something like "/bigsdbs/" (which is the default), or "/cgi-bin/" or just "/" if you don't care about this check.

## 18.3 Administration

### 1. How can I make some isolates public but not others?

The easiest way to do this is to set up two or more separate configuration directories that refer to the database. The URLs to access these will differ by the value of the 'db' attribute, which refers to the name of the configuration directory (in /etc/bigsdbs/dbases/). The database view accessed by each of these configurations can be different as can the access restrictions.

*Example:*

We have a database 'bigsdbs\_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 bigsdbs_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/bigsdbs/dbases/test\_private. This will be accessible from [http://IP\\_ADDRESS/cgi-bin/bigsdbs/bigsdbs.pl?db=test\\_private](http://IP_ADDRESS/cgi-bin/bigsdbs/bigsdbs.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/bigsdbs/dbases/test\_public. This will be accessible from [http://IP\\_ADDRESS/cgi-bin/bigsdbs/bigsdbs.pl?db=test\\_public](http://IP_ADDRESS/cgi-bin/bigsdbs/bigsdbs.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](#)*.





### 19.1 Query operators

Various query forms have operators for use with field values. Available operators are:

- =
  - Exact match (case-insensitive).
- contains
  - Match to a partial string (case-insensitive), e.g. searching for clonal complex ‘contains’ st-11 would return all STs belonging to the ST-11 complex.
- starts with
  - Match to values that start with the search term (case-insensitive).
- ends with
  - Match to values that end with the search term (case-sensitive).
- >
  - Greater than the search term.
- >=
  - Greater than or equal the search term.
- <
  - Less than the search term.
- <=
  - Less than or equal the search term.
- NOT
  - Match to values that do not equal the search term (case-insensitive).

- NOT contain
  - Match to values that do not contain the search term (case-insensitive).

## 19.2 Sequence tag flags

Sequences tagged in the sequence bin can have features indicated by specific flags. The presence of these flags can be queried. These are a superset of *flags available for allele sequences*. Available flags are:

- ambiguous read
  - Genome sequence contains ambiguous nucleotides in coding sequence.
- apparent misassembly
  - Sequence has a region of very high identity to existing allele in one region but looks completely different in another.
- atypical
  - Catch-all term for a sequence that is unusual compared to other alleles of locus.
- contains IS element
  - Coding sequence is interrupted by insertion sequence.
- downstream fusion
  - No stop codon present resulting in translation continuing.
- frameshift
  - Frameshift in sequence relative to other alleles, not resulting in internal stop codon.
- indel
  - Insertion/deletion in sequence that is uncommon compared to other alleles.
- internal stop codon
  - Frameshift in sequence relative to other alleles, resulting in internal stop codon.
- no start codon
  - No apparent start codon in immediate vicinity of usual start.
- no stop codon
  - No stop codon in immediate vicinity of usual stop.
- phase variable: off
  - Coding sequence has a homopolymeric run with a frameshift resulting in a stop codon preventing complete translation.
- truncated
  - Coding sequence is unusually short resulting in a truncated protein (not the same as running off the end of a contig).
- upstream fusion
  - No apparent start codon in immediate vicinity of usual start, likely due to a gene fusion (sequence is transcribed together with upstream coding sequence).

## 19.3 Allele sequence flags

Sequences can be flagged with specific attributes - these are searchable when doing a sequence attribute query. These are a subset of *flags available for tagged sequences*. These are mainly for use with whole genome MLST type data. Multiple flags can be selected by Ctrl-clicking the list. Available flags are:

- atypical
  - Catch-all term for a sequence that is unusual compared to other alleles of locus.
- contains IS element
  - Coding sequence is interrupted by insertion sequence.
- downstream fusion
  - No stop codon present resulting in translation continuing.
- frameshift
  - Frameshift in sequence relative to other alleles, not resulting in internal stop codon.
- indel
  - Insertion/deletion in sequence that is uncommon compared to other alleles.
- internal stop codon
  - Frameshift in sequence relative to other alleles, resulting in internal stop codon.
- no start codon
  - No apparent start codon in immediate vicinity of usual start.
- no stop codon
  - No stop codon in immediate vicinity of usual stop.
- phase variable: off
  - Coding sequence has a homopolymeric run with a frameshift resulting in a stop codon preventing complete translation.
- truncated
  - Coding sequence is unusually short resulting in a truncated protein (not the same as running off the end of a contig).
- upstream fusion
  - No apparent start codon in immediate vicinity of usual start, likely due to a gene fusion (sequence is transcribed together with upstream coding sequence).



## CHAPTER 20

---

### Database schema

---

- Sequence definition database
- Isolate database



## A

### access

- control lists, 44
- restricting, 44

### adding

- classification groups, 84
- isolates, 133
- locus, 53, 55, 57, 62, 88
- MLST scheme, 82
- schemes, 65

### allele definition

- records, 209

### allele designations

- count, 245
- query, 244
- status, 247

### allele sequence

- identify, 213

### allele sequences

- adding, 112

### alleles, 3

- list query, 223

### API authentication

- access token, 412
- accessing protected resources, 413
- consumer key, 410
- request token, 410
- session token, 413
- user authorization, 411

### API resources

- add submission correspondence, 408
- create new submission, 405
- DELETE /db/{database}/submissions/{submission\_id}, 407
- DELETE /db/{database}/submissions/{submission\_id}/files/{filename}, 409
- delete submission record, 407
- delete submission supporting file, 409
- download alleles in FASTA format, 388

download allelic profiles in CSV (tab-delimited) format, 393

download contigs in FASTA format, 400

download submission supporting file, 409

GET /, 383

GET /db, 383

GET /db/{database}, 384

GET /db/{database}/classification\_schemes, 384

GET /db/{database}/classification\_schemes/{classification\_scheme\_id}, 384

GET /db/{database}/classification\_schemes/{classification\_scheme\_id}, 385

GET /db/{database}/contigs/{contig\_id}, 401

GET /db/{database}/fields, 402

GET /db/{database}/fields/{field}, 403

GET /db/{database}/isolates, 395

GET /db/{database}/isolates/{isolate\_id}, 395

GET /db/{database}/isolates/{isolate\_id}/allele\_designations, 397

GET /db/{database}/isolates/{isolate\_id}/allele\_designations/{locus}, 397

GET /db/{database}/isolates/{isolate\_id}/allele\_ids, 398

GET /db/{database}/isolates/{isolate\_id}/contigs, 399

GET /db/{database}/isolates/{isolate\_id}/contigs\_fasta, 400

GET /db/{database}/isolates/{isolate\_id}/history, 400

GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele, 398

GET /db/{database}/isolates/{isolate\_id}/schemes/{scheme\_id}/allele, 399

GET /db/{database}/loci, 385

GET /db/{database}/loci/{locus}, 386

GET /db/{database}/loci/{locus}/alleles, 387

GET /db/{database}/loci/{locus}/alleles/{allele\_id}, 388

GET /db/{database}/loci/{locus}/alleles\_fasta, 388

GET /db/{database}/projects, 403

- GET /db/{database}/projects/{project\_id}, 404
  - GET /db/{database}/projects/{project\_id}/isolates, 404
  - GET /db/{database}/schemes, 391
  - GET /db/{database}/schemes/{scheme\_id}, 391
  - GET /db/{database}/schemes/{scheme\_id}/fields/{field}, 392
  - GET /db/{database}/schemes/{scheme\_id}/loci, 392
  - GET /db/{database}/schemes/{scheme\_id}/profiles, 392
  - GET /db/{database}/schemes/{scheme\_id}/profiles/{profile\_id}, 394
  - GET /db/{database}/schemes/{scheme\_id}/profiles\_csv, 393
  - GET /db/{database}/sequences, 390
  - GET /db/{database}/submissions, 405
  - GET /db/{database}/submissions/{submission\_id}, 406
  - GET /db/{database}/submissions/{submission\_id}/files, 408
  - GET /db/{database}/submissions/{submission\_id}/files/{file\_name}, 409
  - GET /db/{database}/submissions/{submission\_id}/messages, 408
  - GET /db/{database}/users/{user\_id}, 403
  - get summary of defined sequences, 390
  - list allelic profiles defined for scheme, 392
  - list classification schemes, 384
  - list database resources, 384
  - list loci, 385
  - list schemes, 391
  - list site resources, 383
  - POST /db/{database}/isolates/search, 400
  - POST /db/{database}/loci/{locus}/sequence, 389
  - POST /db/{database}/schemes/{scheme\_id}/sequence, 394
  - POST /db/{database}/sequence, 390
  - POST /db/{database}/submissions, 405
  - POST /db/{database}/submissions/{submission\_id}/files, 409
  - POST /db/{database}/submissions/{submission\_id}/messages, 408
  - query allele sequence, 389
  - query allele sequence without specifying locus, 390
  - query scheme sequences, 394
  - retrieve allele identifiers, 398
  - retrieve classification scheme information and groups, 384, 385
  - retrieve contig record, 401
  - retrieve full allele designation record, 397
  - retrieve full allele information, 388
  - retrieve information about scheme field, 392
  - retrieve isolate record, 395
  - retrieve isolate update history, 400
  - retrieve list of allele designations, 397
  - retrieve list of alleles defined for a locus, 387
  - retrieve list of contigs, 399
  - retrieve list of isolate provenance field descriptions, 402
  - retrieve list of isolate records, 395
  - retrieve list of isolates belonging to a project, 404
  - retrieve list of projects, 403
  - retrieve list of scheme allele identifiers, 399
  - retrieve list of submissions, 405
  - retrieve list of supporting files uploaded for submission, 408
  - retrieve locus record, 386
  - retrieve project information, 404
  - retrieve scheme allele designation records, 398
  - retrieve scheme information, 391
  - retrieve scheme loci, 392
  - retrieve specific allelic profile record, 394
  - retrieve submission correspondence, 408
  - retrieve submission record, 406
  - retrieve user information, 403
  - retrieve values set for a provenance field, 403
  - search isolate database, 400
  - upload submission supporting file, 409
  - auto allele definer, 192
    - stop, 195
  - automated assignment
    - scheme profiles, 83
  - autotagger, 189
    - stop, 195
- ## B
- batch profile definitions
    - query, 233
  - batch uploading contigs - multiple isolates, 153
  - BLAST, 314
  - breakdown
    - provenance field, 289
    - sequence bin, 303
    - two-field, 300
  - browse
    - scheme profiles, 225
  - BURST, 319
- ## C
- caching
    - schemes, 47
  - classification groups, 4
    - adding, 84
  - client authorization
    - RESTful interface, 108
  - client databases, 74
  - clustering



- core genome, 87
- codon usage, 323
  - Locus Explorer, 287
- composite fields, 98
- configuration
  - export, 106
- configuration settings
  - validation, 105
- core genome
  - clustering, 87
- count
  - allele designations, 245
  - sequence tags, 248

## D

- defining
  - exemplar alleles, 191

## E

- exemplar alleles
  - defining, 191
- export
  - configuration, 106
- extended attributes
  - locus, 63
  - provenance fields, 101

## F

- filters, 252

## G

- gene presence, 330
- Genome Comparator, 306
- genome filtering, 91
  - in silico hybridization, 94
  - in silico PCR, 91
- GrapeTree
  - minimum-spanning trees, 336
- groups
  - scheme, 71

## H

- hosts
  - mapping, 47

## I

- identify
  - allele sequence, 213
  - sequence type, 215
- in silico hybridization, 94
- in silico PCR, 91
- isolate
  - records, 205

- isolate aliases, 147
- isolate record
  - options, 262
- isolates
  - adding, 133

## L

- linking remote contigs, 157
- list query
  - alleles, 223
- loci, 3
  - modifying display, 267
- locus
  - adding, 53, 55, 57, 62, 88
  - copying existing record, 60
  - extended attributes, 63
- Locus Explorer, 283
  - codon usage, 287
  - polymorphic sites, 283
  - translated sequences, 288
- locus positions
  - setting, 95

## M

- main results table
  - options, 261
- mapping
  - hosts, 47
- minimum-spanning trees
  - GrapeTree, 336
- MLST, 4
- MLST scheme
  - adding, 82
- mod\_perl, 47
- modifying display
  - loci, 267
  - schemes, 267

## O

- offline curation
  - auto allele definer, 192
  - autotagger, 189
- options, 259
  - isolate record, 262
  - main results table, 261
  - provenance fields, 264
  - query, 265

## P

- partitioning
  - sets, 49
- passwords
  - setting, 44

- first user, [45](#)
- performance
  - caching schemes, [47](#)
  - mod\_perl, [47](#)
- permissions, [39](#)
  - locus curation, [42](#)
  - scheme curation, [42](#)
- plugins
  - enabling, [45](#)
- polymorphic sites
  - Locus Explorer, [283](#)
- polymorphisms, [328](#)
- profile
  - records, [211](#)
- profiles, [4](#)
- projects, [162](#)
- provenance field
  - breakdown, [289](#)
- provenance fields
  - extended attributes, [101](#)
  - options, [264](#)
- publications, [148](#)

## Q

- query
  - allele designations, [244](#)
  - batch profile definitions, [233](#)
  - options, [265](#)
  - scheme profiles, [226](#)
  - sequence tags, [249](#)
  - ST definitions from allelic profiles, [230](#)

## R

- records
  - allele definition, [209](#)
  - isolate, [205](#)
  - profile, [211](#)
  - sequence bin, [212](#)
  - sequence tag, [210](#)
- RESTful interface
  - client authorization, [108](#)
- rule-based queries, [80](#)

## S

- scheme
  - groups, [71](#)
- scheme profiles
  - automated assignment, [83](#)
  - browse, [225](#)
  - query, [226](#)
- schemes, [4](#)
  - adding, [65](#)
  - caching, [47](#)

- modifying display, [267](#)
- sequence bin
  - breakdown, [303](#)
  - records, [212](#)
- sequence similarity
  - determining, [236](#)
- sequence tag
  - records, [210](#)
- sequence tags, [4](#)
  - count, [248](#)
  - query, [249](#)
- sequence type
  - identify, [215](#)
- sets, [5](#)
  - partitioning, [49](#)
- ST definitions from allelic profiles
  - query, [230](#)
- stop
  - auto allele definer, [195](#)
  - autotagger, [195](#)

## T

- translated sequences
  - Locus Explorer, [288](#)
- two-field
  - breakdown, [300](#)

## U

- unique combinations, [326](#)
- updates
  - disabling, [46](#)
- uploading contigs - single isolate, [150](#)
- user groups, [39](#)
- user projects, [269](#)
- user types, [39](#)
- users
  - adding, [111](#)