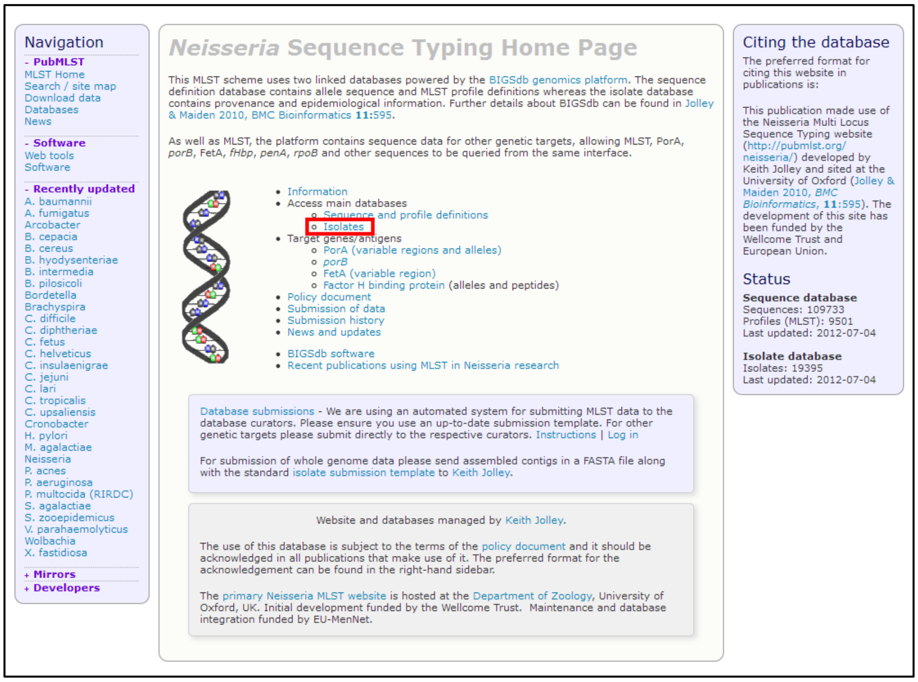
Neighbor-Net analysis using BIGSdb as implemented on PubMLST.org

The whole genome Neighbor-Net analysis described in this paper can be replicated with the following steps:

1. Navigate to the Neisseria database on PubMLST.org (http://pubmlst.org/neisseria/).
2. Select the isolate database:



Neisseria Sequence Typing Home Page

This MLST scheme uses two linked databases powered by the BIGSdb genomics platform. The sequence definition database contains allele sequence and MLST profile definitions whereas the isolate database contains provenance and epidemiological information. Further details about BIGSdb can be found in Jolley & Maiden 2010, BMC Bioinformatics **11**:595.

As well as MLST, the platform contains sequence data for other genetic targets, allowing MLST, PorA, *porB*, FetA, *fHbp*, *penA*, *rpoB* and other sequences to be queried from the same interface.

- Information
- Access main databases
    - Sequence and profile definitions
    - Isolates
- Target genes/antigens
    - PorA (variable regions and alleles)
    - *porB*
    - FetA (variable region)
    - Factor H binding protein (alleles and peptides)
- Policy document
- Submission of data
- Submission history
- News and updates

- BIGSdb software
- Recent publications using MLST in Neisseria research

Database submissions - We are using an automated system for submitting MLST data to the database curators. Please ensure you use an up-to-date submission template. For other genetic targets please submit directly to the respective curators. Instructions | Log in

For submission of whole genome data please send assembled contigs in a FASTA file along with the standard isolate submission template to Keith Jolley.

Website and databases managed by Keith Jolley.

The use of this database is subject to the terms of the policy document and it should be acknowledged in all publications that make use of it. The preferred format for the acknowledgement can be found in the right-hand sidebar.

The primary Neisseria MLST website is hosted at the Department of Zoology, University of Oxford, UK. Initial development funded by the Wellcome Trust. Maintenance and database integration funded by EU-MenNet.

Citing the database

The preferred format for citing this website in publications is:

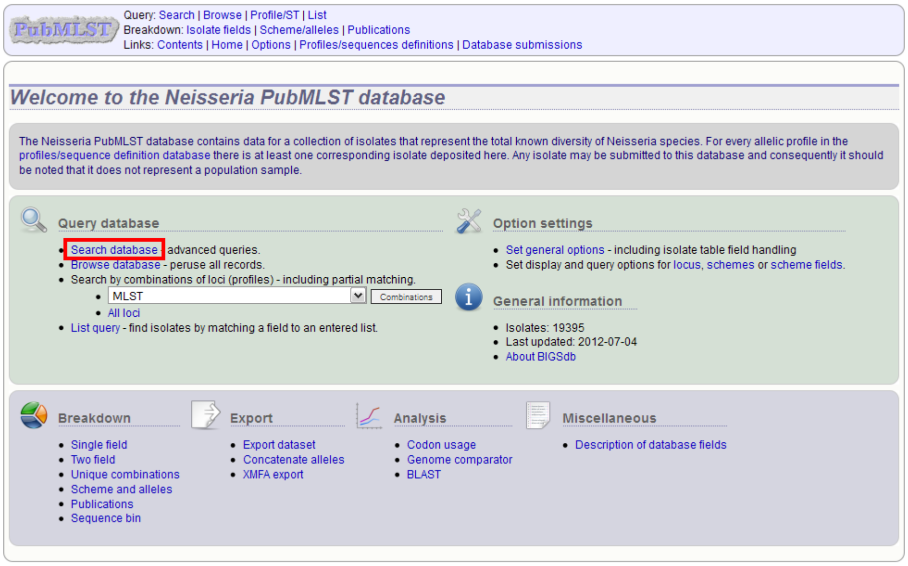
This publication made use of the Neisseria Multi Locus Sequence Typing website (http://pubmlst.org/neisseria/) developed by Keith Jolley and sited at the University of Oxford (Jolley & Maiden 2010, *BMC Bioinformatics*, **11**:595). The development of this site has been funded by the Wellcome Trust and European Union.

Status

Sequence database
Sequences: 109733
Profiles (MLST): 9501
Last updated: 2012-07-04

Isolate database
Isolates: 19395
Last updated: 2012-07-04

3. Click ‘Search database’:



Welcome to the 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.
    - MLST Combinations
    - All loci
- List query - find isolates by matching a field to an entered list.

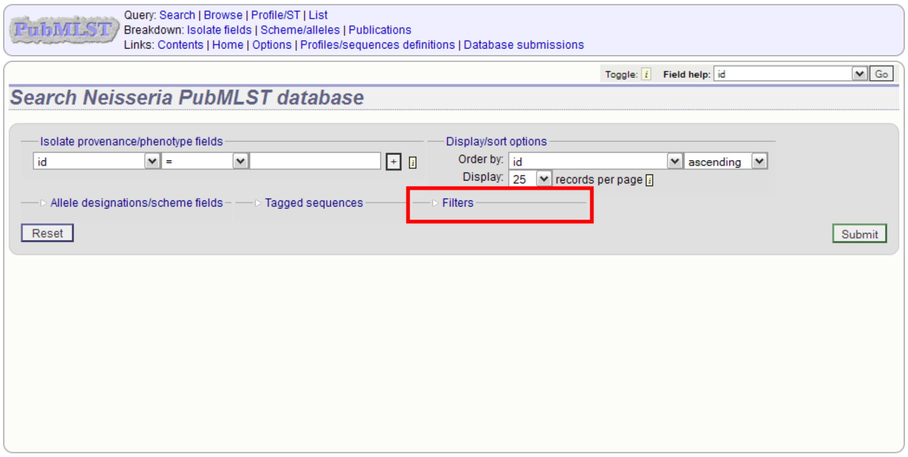
Option settings

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

General information

- Isolates: 19395
- Last updated: 2012-07-04
- About BIGSdb

4. Click the filters list:

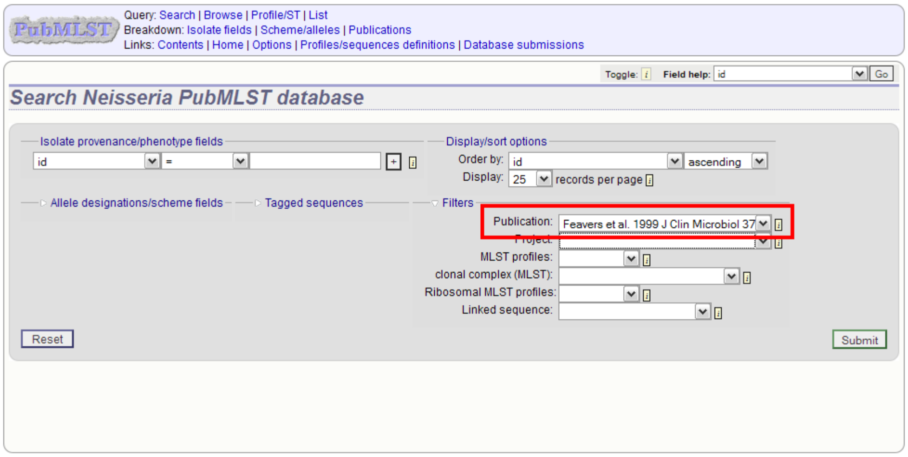


id =

Order by: id ascending

Display: 25 records per page

5. Select the Feavers *et al.* 1999 paper from the publication filter list and click 'Submit':



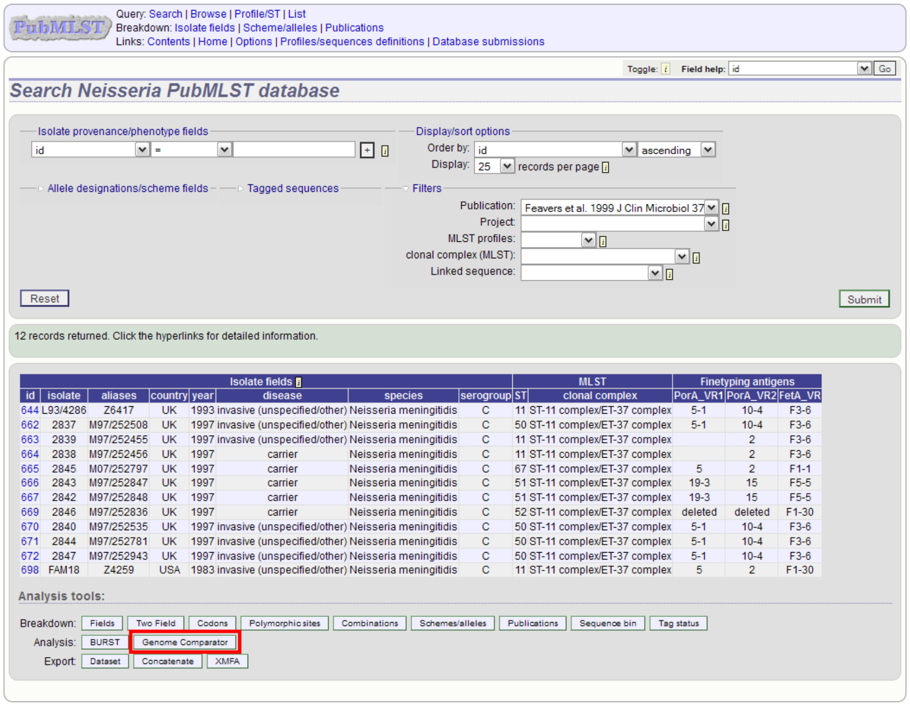
id =

Order by: id ascending

Display: 25 records per page

Publication: Feavers et al. 1999 J Clin Microbiol 37

6. The records for isolates from the outbreak will be displayed. Click the ‘Genome Comparator’ button at the bottom of the table:



Search Neisseria PubMLST database

Field help: id

id =

Order by: id ascending

Display: 25 records per page

Publication: Feavers et al. 1999 J Clin Microbiol 37

Project:

MLST profiles:

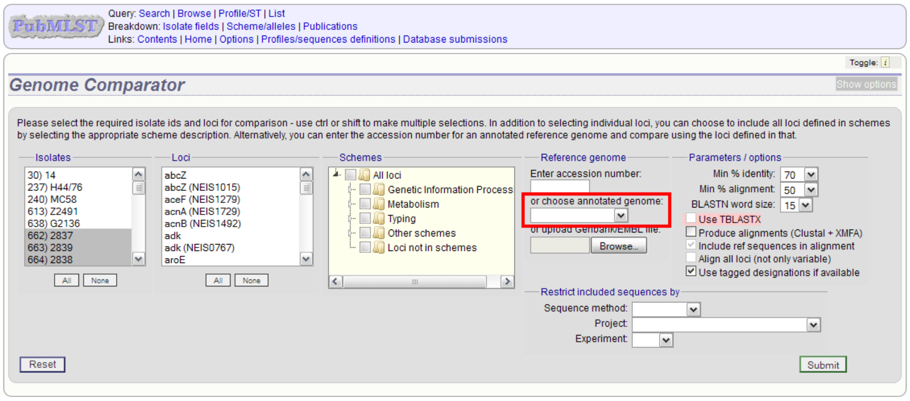
clonal complex (MLST):

Linked sequence:

12 records returned. Click the hyperlinks for detailed information.

Isolate fields								MLST		Finetyping antigens		
id	isolate	aliases	country	year	disease	species	serogroup	ST	clonal complex	PorA_VR1	PorA_VR2	FetA_VR
644	L93/4286	Z6417	UK	1993	invasive (unspecified/other)	Neisseria meningitidis	C	11	ST-11 complex/ET-37 complex	5-1	10-4	F3-6
662	2837	M97/252508	UK	1997	invasive (unspecified/other)	Neisseria meningitidis	C	50	ST-11 complex/ET-37 complex	5-1	10-4	F3-6
663	2839	M97/252455	UK	1997	invasive (unspecified/other)	Neisseria meningitidis	C	11	ST-11 complex/ET-37 complex		2	F3-6
664	2838	M97/252456	UK	1997	carrier	Neisseria meningitidis	C	11	ST-11 complex/ET-37 complex		2	F3-6
665	2845	M07/252797	UK	1997	carrier	Neisseria meningitidis	C	67	ST-11 complex/ET-37 complex	5	2	F1-1
666	2843	M97/252847	UK	1997	carrier	Neisseria meningitidis	C	51	ST-11 complex/ET-37 complex	19-3	15	F5-5
667	2842	M97/252848	UK	1997	carrier	Neisseria meningitidis	C	51	ST-11 complex/ET-37 complex	19-3	15	F5-5
669	2846	M97/252836	UK	1997	carrier	Neisseria meningitidis	C	52	ST-11 complex/ET-37 complex	deleted	deleted	F1-30
670	2840	M97/252535	UK	1997	invasive (unspecified/other)	Neisseria meningitidis	C	50	ST-11 complex/ET-37 complex	5-1	10-4	F3-6
671	2844	M97/252781	UK	1997	invasive (unspecified/other)	Neisseria meningitidis	C	50	ST-11 complex/ET-37 complex	5-1	10-4	F3-6
672	2847	M97/252943	UK	1997	invasive (unspecified/other)	Neisseria meningitidis	C	50	ST-11 complex/ET-37 complex	5-1	10-4	F3-6
698	FAM18	Z4259	USA	1983	invasive (unspecified/other)	Neisseria meningitidis	C	11	ST-11 complex/ET-37 complex	5	2	F1-30

7. The isolates will be pre-selected. Select ‘FAM18’ from the ‘annotated genome’ drop down list box:



Genome Comparator

Please select the required isolate ids and loci for comparison - use ctrl or shift to make multiple selections. In addition to selecting individual loci, you can choose to include all loci defined in schemes by selecting the appropriate scheme description. Alternatively, you can enter the accession number for an annotated reference genome and compare using the loci defined in that.

Isolates:
30) 14
237) H44/76
240) MC58
613) Z2491
638) G2136
662) 2837
663) 2839
664) 2838

Loci:
abcZ
abcZ (NEIS1015)
aceF (NEIS1279)
acnA (NEIS1729)
acnB (NEIS1492)
adk
adk (NEIS0767)
aroE

Schemes:
☐ All loci
- ☐ Genetic Information Process
- ☐ Metabolism
- ☐ Typing
- ☐ Other schemes
- ☐ Loci not in schemes

Enter accession number:

or choose annotated genome:

or upload GenBank/EMBL file:

Min % identity: 70

Min % alignment: 50

BLASTN word size: 15

☐ Use TBLASTX

☐ Produce alignments (Clustal + XMFA)

☑ Include ref sequences in alignment

☐ Align all loci (not only variable)

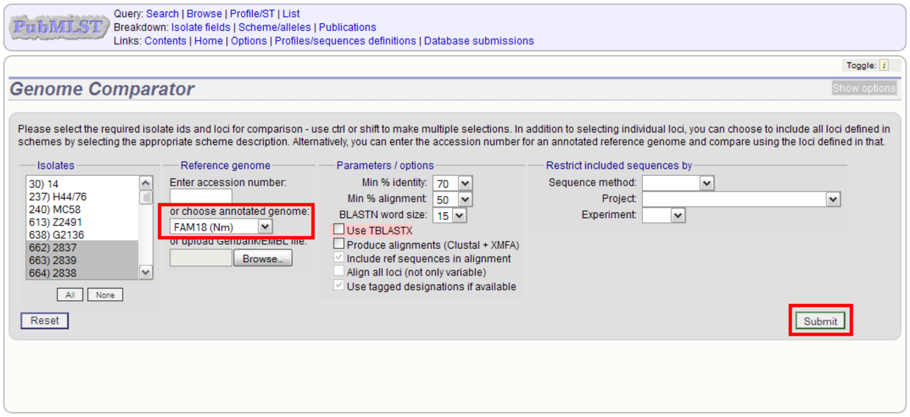
☑ Use tagged designations if available

Sequence method:

Project:

Experiment:

8. The locus and scheme lists will be hidden when a genome is selected. Press ‘Submit’.



Genome Comparator

Please select the required isolate ids and loci for comparison - use ctrl or shift to make multiple selections. In addition to selecting individual loci, you can choose to include all loci defined in schemes by selecting the appropriate scheme description. Alternatively, you can enter the accession number for an annotated reference genome and compare using the loci defined in that.

Isolates
30) 14
237) H44/76
240) MC58
613) Z2491
638) G2136
662) 2837
663) 2839
664) 2838

Reference genome
Enter accession number:
or choose annotated genome: FAM18 (Nm)
or upload Genbank/EMBL file:

Parameters / options
Min % identity: 70
Min % alignment: 50
BLASTN word size: 15
☐ Use TBLASTX
☐ Produce alignments (Clustal + XMFA)
☑ Include ref sequences in alignment
☐ Align all loci (not only variable)
☑ Use tagged designations if available

Restrict included sequences by
Sequence method:
Project:
Experiment:

9. The analysis job will be submitted to the server job queue. Click the job hyperlink to display the job viewer page. This will auto-refresh until the analysis is completed:

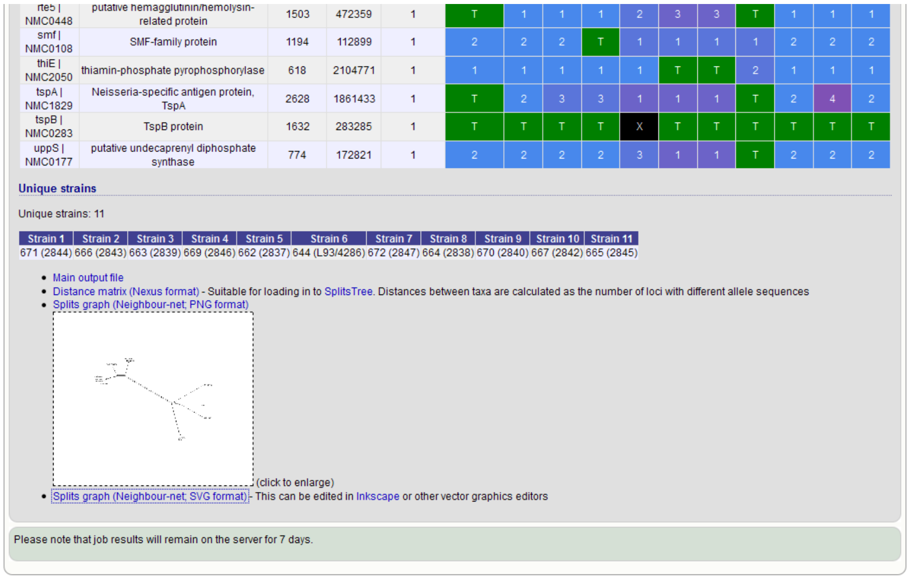
Genome Comparator

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.

10. A Neighbor-Net will be generated and displayed at the bottom of the page. Various output files are hyperlinked including the distance matrix used to generate the network. This can be loaded in to stand-alone Splitstree for further manipulation.



rte5 | NMC0448	putative hemagglutinin/hemolysin-related protein	1503	472359	1	T	1	1	1	2	3	3	T	1	1	1
smf | NMC0108	SMF-family protein	1194	112899	1	2	2	2	T	1	1	1	1	2	2	2
thiE | NMC2050	thiamin-phosphate pyrophosphorylase	618	2104771	1	1	1	1	1	1	T	T	2	1	1	1
tspA | NMC1829	Neisseria-specific antigen protein, TspA	2628	1861433	1	T	2	3	3	1	1	1	T	2	4	2
tspB | NMC0283	TspB protein	1632	283285	1	T	T	T	T	X	T	T	T	T	T	T
uppS | NMC0177	putative undecaprenyl diphosphate synthase	774	172821	1	2	2	2	2	3	1	1	T	2	2	2

Unique strains

Unique strains: 11

Strain 1	Strain 2	Strain 3	Strain 4	Strain 5	Strain 6	Strain 7	Strain 8	Strain 9	Strain 10	Strain 11
671 (2844)	666 (2843)	663 (2839)	669 (2846)	662 (2837)	644 (L93/4286)	672 (2847)	664 (2838)	670 (2840)	667 (2842)	665 (2845)

- Main output file
- 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

Please note that job results will remain on the server for 7 days.