PubMLST and the BIGSdb genomics platform

Keith Jolley

UNIVERSITY OF OXFORD

PubMLST

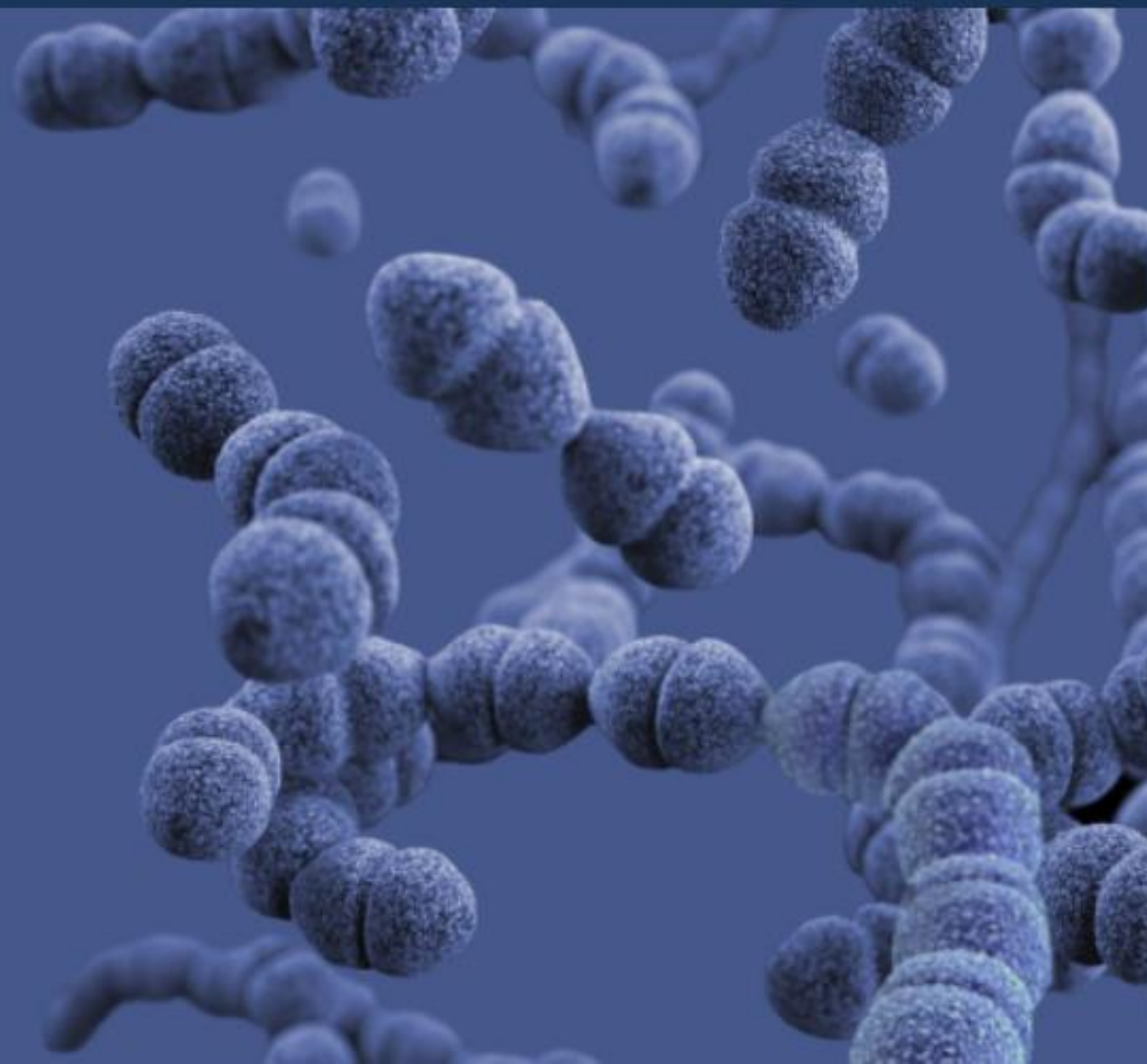
Public databases for molecular typing and microbial genome diversity

MY ACCOUNT

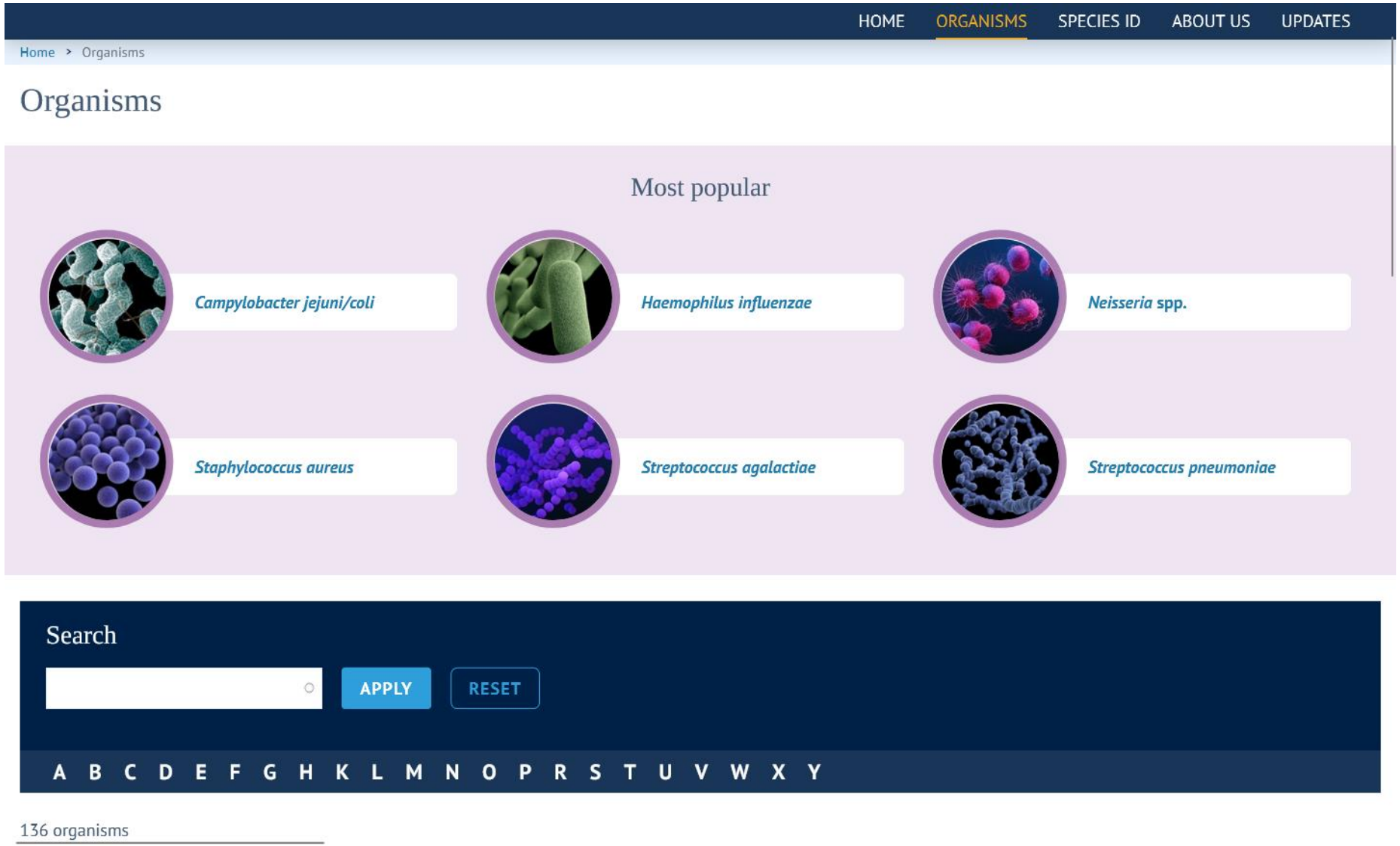
HOME ORGANISMS SPECIES ID ABOUT US UPDATES

A collection of open-access, curated databases that integrate population sequence data with provenance and phenotype information for over 130 different microbial species and genera.

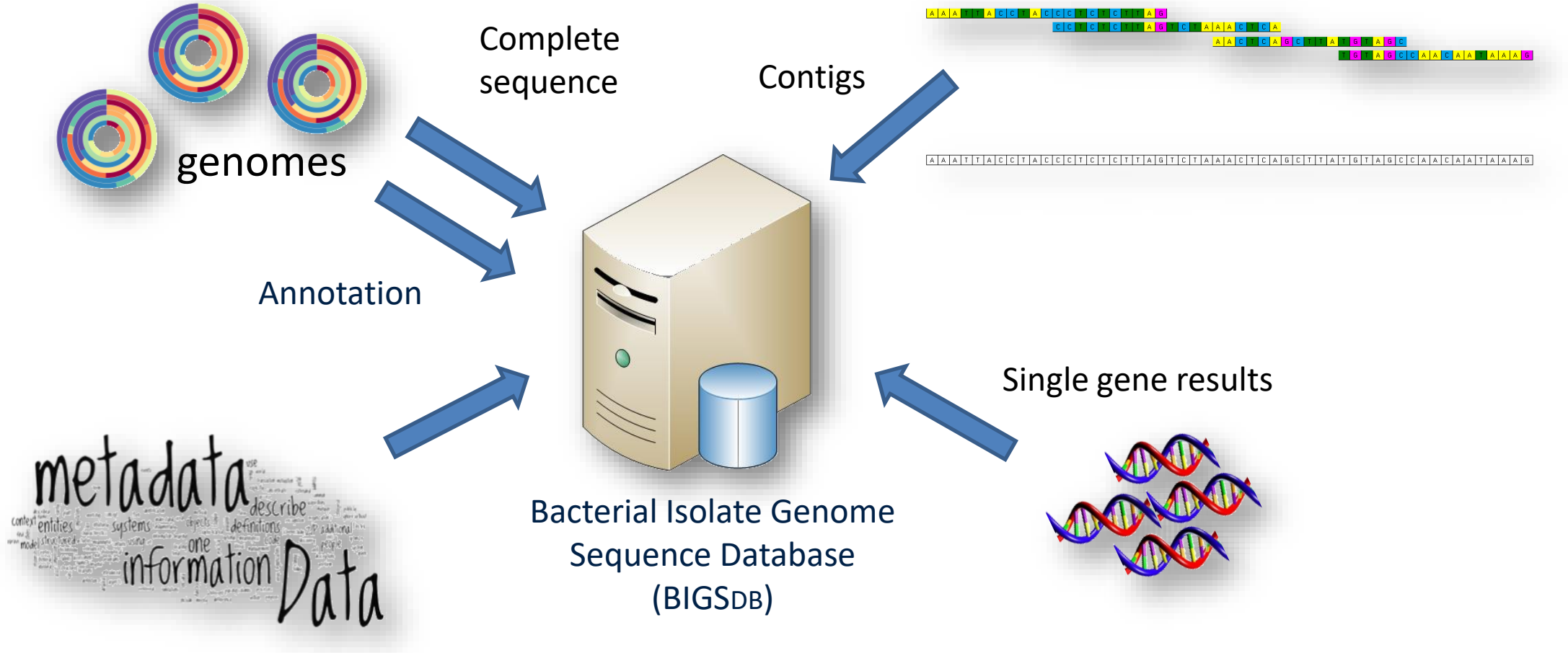
36,103,520	1,275,518	999,279
ALLELES	ISOLATES	GENOMES



PubMLST hosts typing nomenclatures for >130 microorganisms

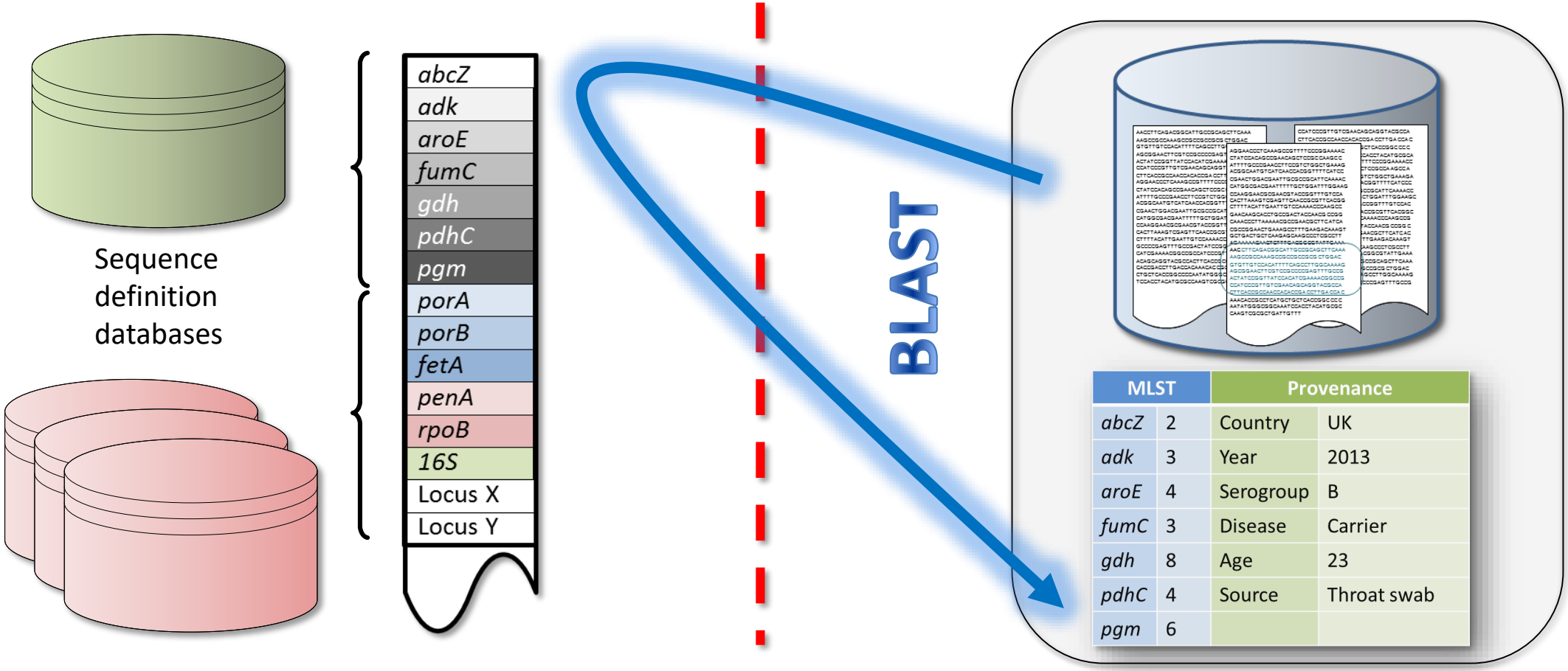


Population genomics: the BIGSdb platform



Jolley & Maiden 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11:**595

Bacterial Isolate Genome Sequence Database (BIGSdb) design philosophy



Jolley & Maiden 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**:595

BIGSdb consists of two main database structures

Sequence definitions


```
Allele1: TTTGATACTGTTGCCGAAGGTTTCCC
Allele2: TTTGATACCGTTGCCGAAGGTTTCCC
Allele3: TTTGATTCCGTTGCCGAAGGTTTCCC
Allele4: TTTGATTCCGATGCCGAAGGTTTCCC
```

Allelic profiles

ST	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	clonal_complex
1	1	3	1	1	1	1	3	ST-1 complex
2	1	3	4	7	1	1	3	ST-1 complex
3	1	3	1	1	1	23	13	ST-1 complex
4	1	3	3	1	4	2	3	ST-4 complex
5	1	1	2	1	3	2	3	ST-5 complex
6	1	1	2	1	3	2	11	ST-5 complex
7	1	1	2	1	3	2	19	ST-5 complex
8	2	3	7	2	8	5	2	ST-8 complex
9	2	3	8	10	8	5	2	ST-8 complex
10	2	3	4	2	8	15	2	ST-8 complex
11	2	3	4	3	8	4	6	ST-11 complex
12	4	3	2	16	8	11	20	

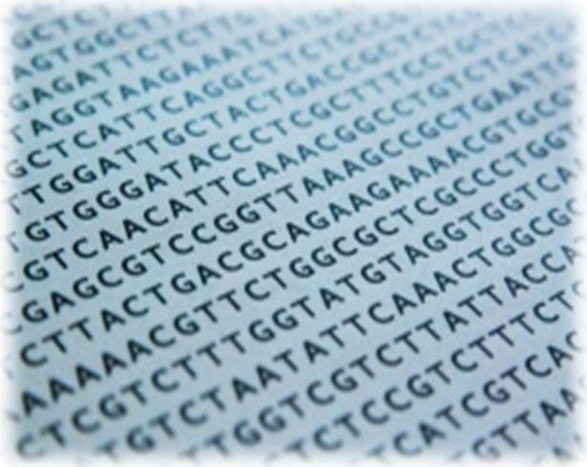
Nomenclature

BIGSdb consists of two main database structures

Provenance

Phenotype

Sequences



Isolate data

BIGSdb as a nomenclature server

BIGSdb is a nomenclature server

PubMLST Public databases for molecular typing and microbial genome diversity

MY ACCOUNT

HOME ORGANISMS SPECIES ID ABOUT US UPDATES

Home > Organisms > *Neisseria* spp. > *Neisseria* typing > Sequence query

Help

Sequence query

Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences.

Please select locus/scheme: MLST

Order results by: locus

Enter query sequence (single or multiple contigs up to whole genome in size)

```
>187372 NODE_1118_length_222_cov_28.878378
TCCCTGTGGGAGAGGGCTAGGGAGAGGGCGGCAAACCGCAGGTTTGCTTGGGCGGCATTT
TCAACGTGCAGGCTGCTTTTCGATTTTTTGCAGCTTCGGTTTTAGCTTCGCAGAAACTCT
GCTTCCTTCGAAAGCTCCGTTTTCAGACGACCTTTCAGTTTTCAGTGCACACAGGCTTTG
TGAAAGCGTTATCGGCTTTGATGTAAGCCTACGGCTTACTTTCCCCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCCGGG
>187373 NODE_320_length_7232_cov_48.889935
```

Alternatively upload FASTA file

Select FASTA file: Browse... No file selected.

or enter Genbank accession

Action

RESET SUBMIT

Contact

Get in touch with us if you have any comments or suggestions concerning the website and the databases.

Follow

Supported by

wellcome

Disclaimer & Privacy | Cookies | Terms & Conditions

Website by Manta Ray Media

BIGSdb is a nomenclature server

7 exact matches found.

Locus	Allele	Length	Contig	Start position	End position	Flags	Comments
abcZ	2	433	187414	5637	6069		
adk	3	465	187432	8935	9399		
aroE	4	490	187449	4826	5315		
fumC	3	465	187396	4871	5335		
gdh	8	501	187466	7333	7833		
pdhC	4	480	187542	56831	57310		
pgm	6	450	187395	21997	22446		

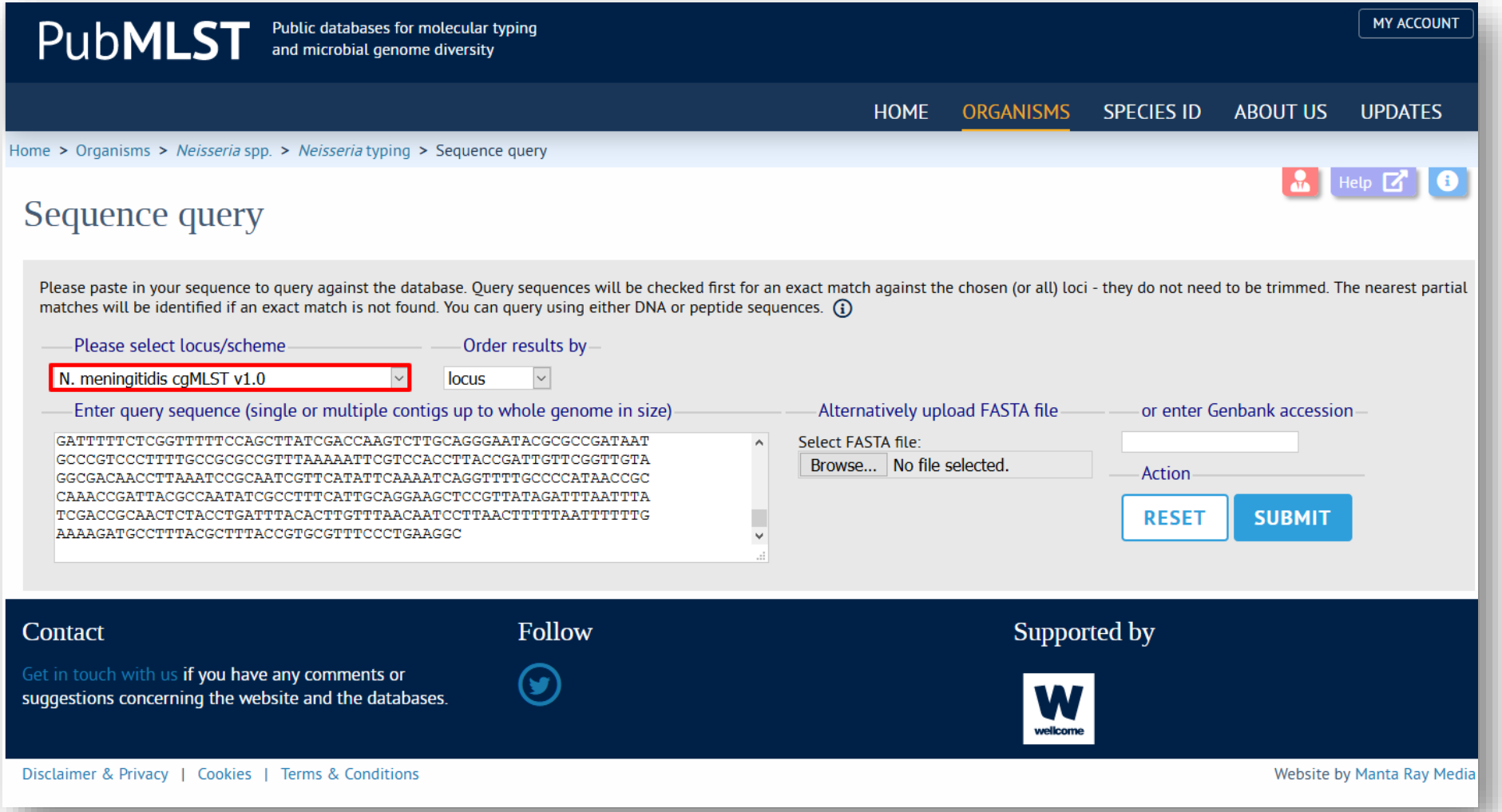
Only exact matches are shown above. If a locus does not have an exact match, try querying specifically against that locus to find the closest match.

MLST

Matching profile

ST: 11
clonal complex: ST-11 complex

Closest matching genomes can be identified by cgMLST clustering



Closest matching genomes can be identified by cgMLST clustering

N. meningitidis cgMLST v1.0

Matching profiles

Closest profile: cgST-56
Mismatches: 5
Loci matched: 1600/1605 (99.7%)

Similar profiles (determined by classification schemes)

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

Classification scheme	Clustering method	Mismatch threshold	Status	Group	Profiles	Isolates
Nm_cgc_200	Single-linkage	200	experimental	65	4810	PubMLST isolates 3474
Nm_cgc_100	Single-linkage	100	experimental	42	1	PubMLST isolates 1
Nm_cgc_50	Single-linkage	50	experimental	50	1	PubMLST isolates 1
Nm_cgc_25	Single-linkage	25	experimental	50	1	PubMLST isolates 1

BIGSdb hosts annotated isolate and genome data

Annotation links genomes with provenance and phenotype

HOME ORGANISMS SPECIES ID ABOUT US UPDATES

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

Help

Full information on isolate 7891 (id:7)

Projects

Projects

This isolate is a member of the following project:

107 global collection

This dataset was originally used to validate MLST and was chosen to represent global diversity of N. meningitidis in the latter half of the Twentieth Century. It has been used in many publications since and the isolates are available as the EMGM MLST reference collection.

Provenance

Provenance/primary metadata

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
genogroup:	A
genogroup notes:	A backbone: All essential capsule genes intact and present. Prediction code: https://github.com/ntopaz/characterize_neisseria_capsule.
capsule group:	A
MLEE designation:	Subgroup III
serotype:	4,21
sero subtype:	P1.9
ET no:	48
biosample accession:	ERS006946 www.ebi.ac.uk
comments:	Pili I,IIa
sender:	Wendell Zollinger, Dept Bacterial Diseases,Walter Reed Army Institute of Research, Washington DC, USA
curator:	Auto Tagger
update history:	122 updates show details
date entered:	2001-02-07
datestamp:	2020-10-06

Annotation links genomes with provenance and phenotype

Phenotype

Publications

Genome sequence

Secondary metadata

Vaccines

Bexsero reactivity: exact match

notes

Bexsero notes: NadA_peptide: 8 is exact match to vaccine variant - peptide sequence match (PMID:27521232)

Trumenba reactivity: insufficient data

notes

Publications (8)

- 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**
- Bratcher HB, Corton C, Jolley KA, Parkhill J, Maiden MC (2014). A gene-by-gene population genomics platform: de novo assembly, annotation and genealogical analysis of 108 representative Neisseria meningitidis genomes. *BMC Genomics* **15**:1138 **108 isolates**
- Didelot X, Urwin R, Maiden MC, Falush D (2009). Genealogical typing of Neisseria meningitidis. *Microbiology* **155**:3176-86 **93 isolates**

Sequence bin

contigs:	199	**mean length:**	10,339 bp	**N90 contig number:**	94	**N95 length (L95):**	3,616
total length:	2,057,385 bp	**N50 contig number:**	29	**N90 length (L90):**	6,218	**loci tagged:**	2,213
max length:	112,831 bp	**N50 length (L50):**	23,361	**N95 contig number:**	116		

Show sequence bin

Annotation links genomes with provenance and phenotype

Clusters

Annotation

Tools

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: 1 (360 isolates)
Nm_cgc_100	N. meningitidis cgMLST v1.0	Single-linkage	100	experimental	group: 5 (5 isolates)

Schemes and loci

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

Navigate and select schemes within tree to display allele designations

Tools

Analysis: rMLST species id PCR

Contact

Get in touch with us if you have any comments or suggestions concerning the website and the databases.

Follow

Supported by

wellcome

Disclaimer & Privacy | Cookies | Terms & Conditions

Website by Manta Ray Media

Datasets can be searched by wide-range of criteria

HOME ORGANISMS SPECIES ID ABOUT US UPDATES

Home > Organisms > *Neisseria* spp. > *Neisseria* isolates > Search or browse database

Help

Search or browse database

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

Isolate provenance fields

Combine with: AND

country = France

year >= 2010

Filters

Publication: Select options

Project: Select options

MLST profiles: complete

Clonal complex (MLST): ST-41/44 complex

Sequence bin: Sequence bin size >= 2 Mbp

Include old record versions

Add filter: Add

Display/sort options

Order by: id ascending

Display: 25 records per page

Action

RESET SEARCH

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

Your projects

Select project... Add these records

Bookmark query

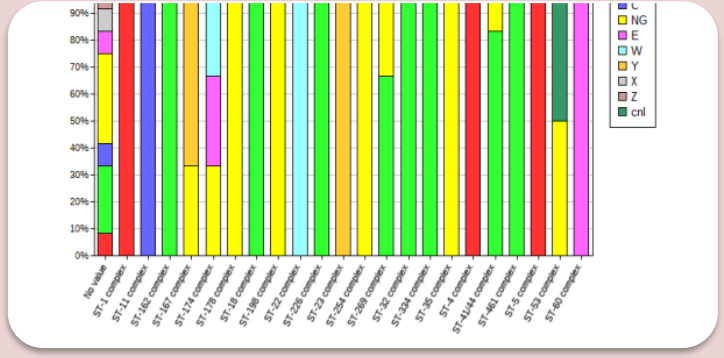
2020-10-19:1 Add bookmark

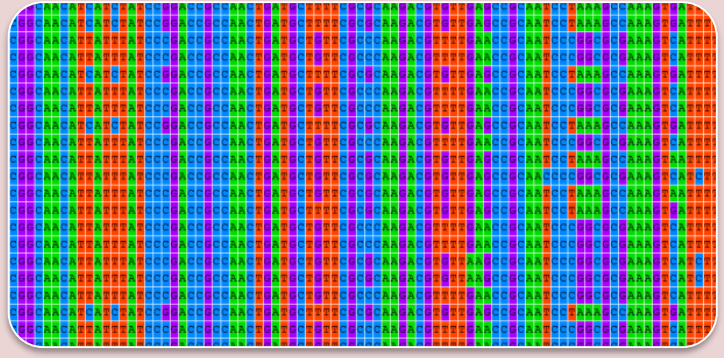
« ‹ 1 2 3 4 › »

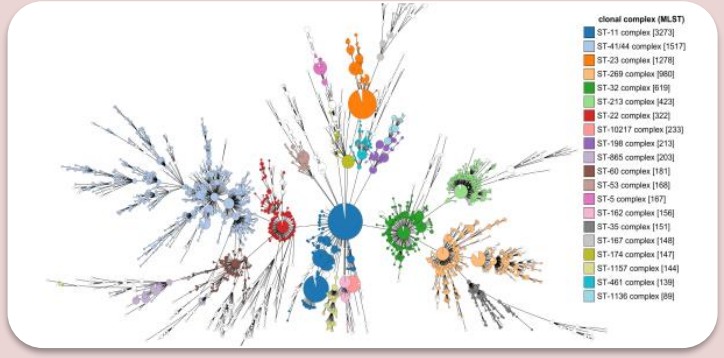
Isolate fields																			Seqbin size (bp)	MLST		Finetyping antigens		
id	isolate	aliases	country	region	year	date sampled	date received	non culture	age yr	age range	age mth	sex	disease	source	species	capsule group	Bexsero reactivity	Trumenba reactivity		ST	clonal complex	PorA_VR1	PorA_VR2	FetA_VR
39963	LNP27359		France		2013										Neisseria meningitidis	B	exact match	insufficient data	2,139,700	1403	ST-41/44 complex	7-2	4	F1-5
39966	LNP27364		France		2013										Neisseria meningitidis	B	insufficient data	cross-reactive	2,102,024	3532	ST-41/44 complex	7-1	1	F1-88
39970	LNP27374		France		2013										Neisseria meningitidis	B	exact match	cross-reactive	2,069,422	1403	ST-41/44 complex	7-2	4	F1-5
39976	LNP27388		France		2013										Neisseria meningitidis	B	exact match	cross-reactive	2,075,331	41	ST-41/44 complex	7-2	4	F1-5
39993	LNP27432		France		2013										Neisseria meningitidis	B	exact match	cross-reactive	2,112,922	41	ST-41/44 complex	7-2	4	F1-5
40009	LNP27478		France		2013										Neisseria meningitidis	B	exact match	cross-reactive	2,071,915	7385	ST-41/44 complex	7-2	4	F1-5
40011	LNP27486		France		2014										Neisseria	B	insufficient	insufficient	2,136,451	10685	ST-41/44	7-1	15-1	F3-6

Export and analysis plugins

Plugins







Breakdown

- Single fields
- Pivot tables (2 fields)
- Combinations
- Polymorphic sites
- Sequence bin

Export

- Isolate dataset
- Contigs
- Aligned sequences

Analysis

- Genome Comparator
- GrapeTree
- PhyloViz
- iTOL
- Microreact
- In silico PCR

Using BIGSdb to compare bacterial genomes

Genomes can be rapidly compared using gene-by-gene variant numbering approach

Locus	Product	Sequence length	Genome position	Reference genome	644 (L93/4286)	662 (2837)	665 (2845)	666 (2843)	667 (2842)	669 (2846)	670 (2840)	671 (2844)	672 (2847)	698 (FAM18)	41784 (2839)	41785 (2838)
NMC0001|lpxC|envA	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosmine deacetylase	924	1261	1	1	1	1	1	1	1	1	1	1	1	1	2
NMC0002|pilS1	pilin (fragment)	291	3341	1	1	1	1	1	1	1	1	1	1	1	2	1
NMC0003|pilS2	truncated pilin	366	3675	1	2	2	2	2	2	2	2	2	2	1	2	2
NMC0004|fbp	peptidyl-prolyl cis-trans isomerase	330	4069	1	2	2	2	2	2	I	2	2	2	1	2	2
NMC0005	putative membrane protein	219	4476	1	2	3	3	3	3	4	3	5	3	1	3	3
NMC0006	putative glycerate dehydrogenase	954	4816	1	2	2	2	2	2	2	2	2	2	1	2	2
NMC0007|metG	methionyl-tRNA synthetase	2058	5843	1	2	2	2	2	2	2	2	2	2	1	3	3
NMC0008|glmS	glucosamine--fructose-6-phosphate aminotransferase [isomerizing]	1839	8016	1	2	2	2	2	2	2	2	2	2	1	2	2
NMC0009	putative lipoprotein	519	10290	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0010|gna33	outer membrane lipoprotein Gna33	1326	11226	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0011	putative integral membrane protein	840	12763	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0012	putative lipoprotein	1167	13599	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0013	possible membrane protein	1266	15029	1	2	1	1	3	3	1	1	1	1	1	1	1
NMC0014|phnA	putative phosphonoacetate hydrolase	330	16366	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0015|glmU	bifunctional GlmU protein [includes: UDP-N-acetylglucosamine pyrophosphorylase (EC 2.7.7.23) (N-acetylglucosamine-1-phosphate uridyltransferase); glucosamine-1-phosphate N-acetyltransferase (EC 2.3.1.57)]	1371	16772	1	1	1	1	1	1	2	1	1	1	1	1	1
NMC0016	conserved hypothetical protein	978	18197	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0017|tbpA	putative solute-binding periplasmic protein	1002	19233	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0018	putative inner membrane protein	849	20328	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0019	conserved hypothetical protein	486	21203	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0020|pilB	peptide methionine sulfoxide reductase	1569	21790	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0021|pilA	probable signal recognition particle protein	1266	23503	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0022	putative transposase	957	25628	1	2	2	1	1	1	1	2	2	2	1	2	2
NMC0023	conserved hypothetical protein (putative ATP-binding protein)	1188	26799	1	I	2	I	3	3	3	2	2	2	1	3	3
NMC0024	putative inner membrane protein	285	27987	1	2	2	3	3	3	3	2	2	2	1	2	2
NMC0025	TspB protein	1560	28574	1	I	I	X	I	I	I	I	I	I	1	2	3
NMC0027	putative inner membrane protein	303	30552	1	2	2	3	2	2	I	2	2	2	1	4	4
NMC0028	putative periplasmic protein	231	30882	1	1	1	2	1	1	1	1	1	1	1	1	1
NMC0029	conserved hypothetical protein	201	31339	1	1	1	1	1	1	1	1	1	1	1	1	1
NMC0030	conserved hypothetical protein	312	31543	1	2	3	1	1	1	I	3	3	3	1	4	2

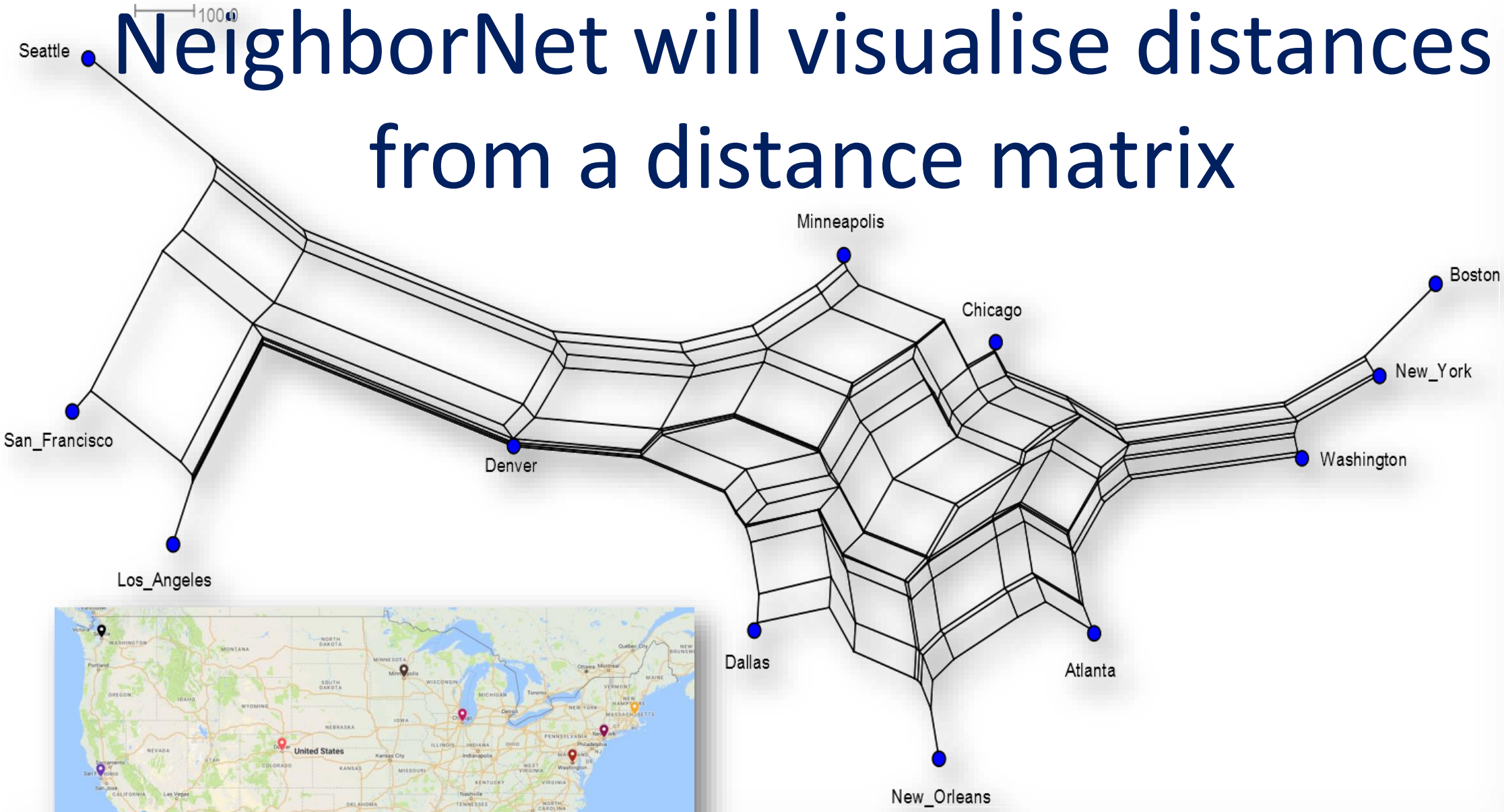
Distance matrix shows number of genes that are different between every pair of isolates

```
#NEXUS
[Distance matrix calculated by BIGSdb Genome Comparator (Thu Sep 12 09:43:44 2013)]
[Jolley & Maiden 2010 BMC Bioinformatics 11:595]
[Truncated loci excluded from analysis]
[Paralogous loci included in analysis]

BEGIN taxa;
   DIMENSIONS ntax = 13;

END;

BEGIN distances;
   DIMENSIONS ntax = 13;
   FORMAT
      triangle=LOWER
      diagonal
      labels
      missing=?
   ;
MATRIX
ref      0
644|L93/4286    626     0
662|2837        651     146     0
663|2839        603     189     221     0
664|2838        604     189     221     14      0
665|2845        366     671     689     645     645     0
666|2843        374     656     673     634     635     446     0
667|2842        379     662     675     639     639     441     28      0
669|2846        257     650     682     634     633     408     421     425     0
670|2840        657     153     33      235     235     696     677     683     692     0
671|2844        650     143     20      227     227     691     672     677     683     31      0
672|2847        649     147     21      223     221     690     670     676     683     34      26      0
698|FAM18       0       626     651     603     604     366     374     379     257     657     650     649     0
   ;
END;
```

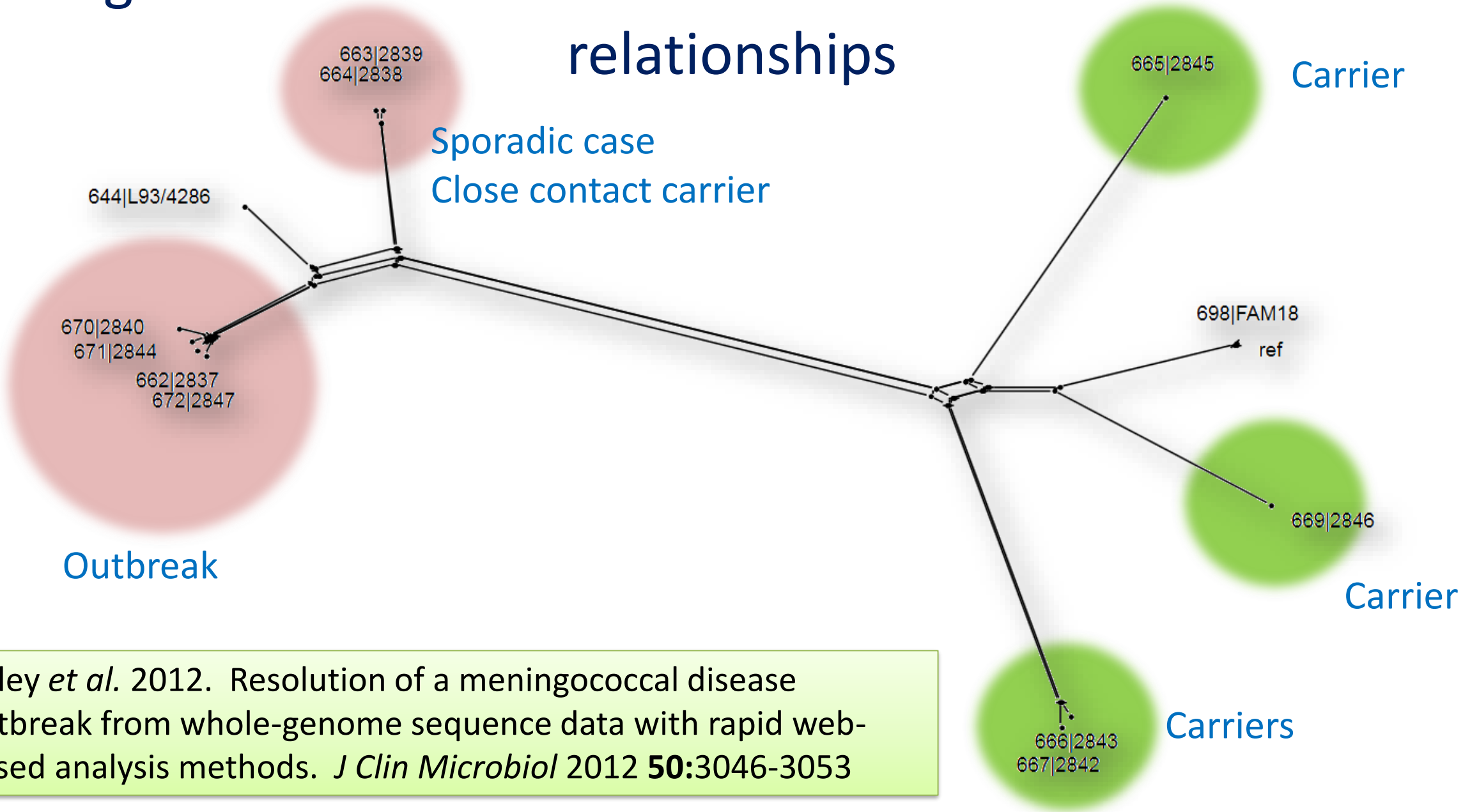
Genetic distances can be treated in the same way as geographic distances

New York	0											
Washington	204	0										
Boston	190	394	0									
Dallas	1372	1183	1551	0								
New Orleans	1170	966	1360	443	0							
Los Angeles	2448	2297	2594	1239	1671	0						
San Francisco	2569	2438	2696	1483	1924	348	0					
Atlanta	747	543	937	720	425	1935	2138	0				
Chicago	712	594	850	806	836	1744	1857	589	0			
Denver	1629	1491	1767	663	1082	831	948	1211	919	0		
Minneapolis	1017	932	1122	865	1054	1523	1583	909	355	699	0	
Seattle	2405	2324	2488	1682	2101	961	680	2181	1735	1021	1393	0

NeighborNet will visualise distances from a distance matrix



NeighborNet visualises distance matrix to resolve strain relationships



Jolley *et al.* 2012. Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid web-based analysis methods. *J Clin Microbiol* 2012 **50:**3046-3053

GrapeTree

GrapeTree analysis can be used to investigate phenotype/provenance compared to genotype