Bacterial Strain Taxonomy for Genomic Surveillance

# MLST and cgMLST: Concepts and Applications

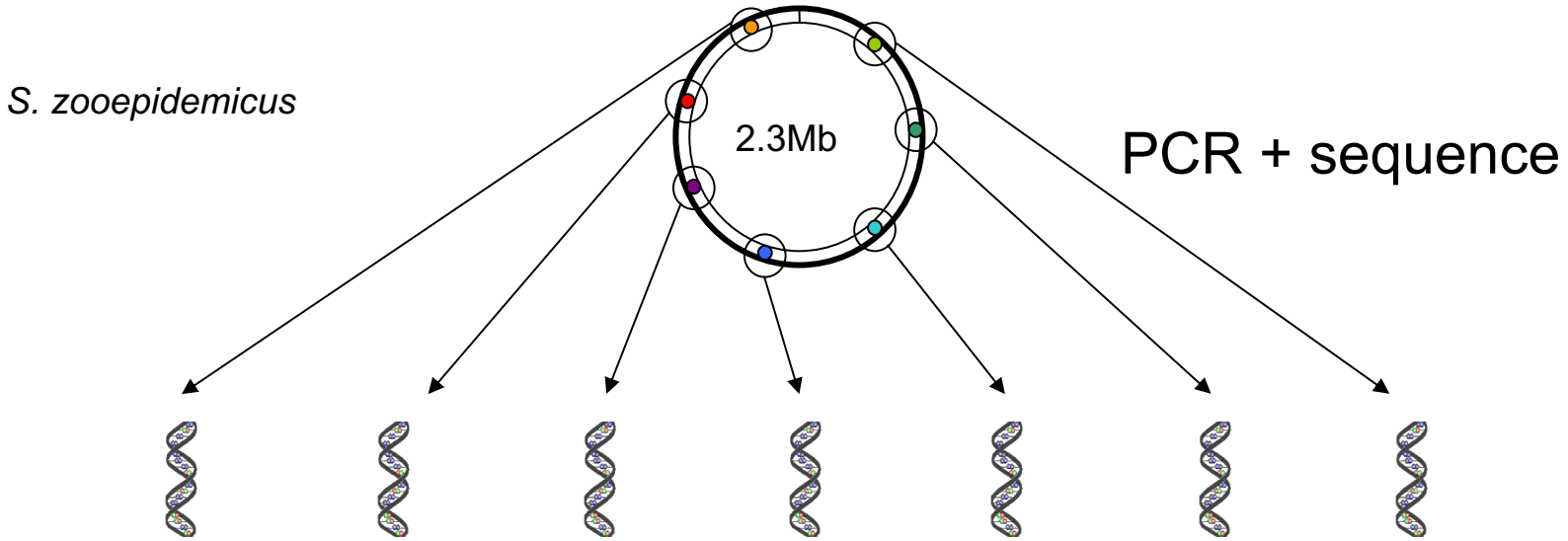
October 2025

# Intended Learning Objectives

Specific objectives of this session:

1. Learn about the principles of MLST and cgMLST
2. Learn about allele calling and scheme design
3. Understand applications in surveillance and outbreak detection
4. Learn about the strengths and limitations of MLST/cgMLST
5. Explore implementations of LIN codes

# The principles of Multi-Locus Sequence Typing (MLST)

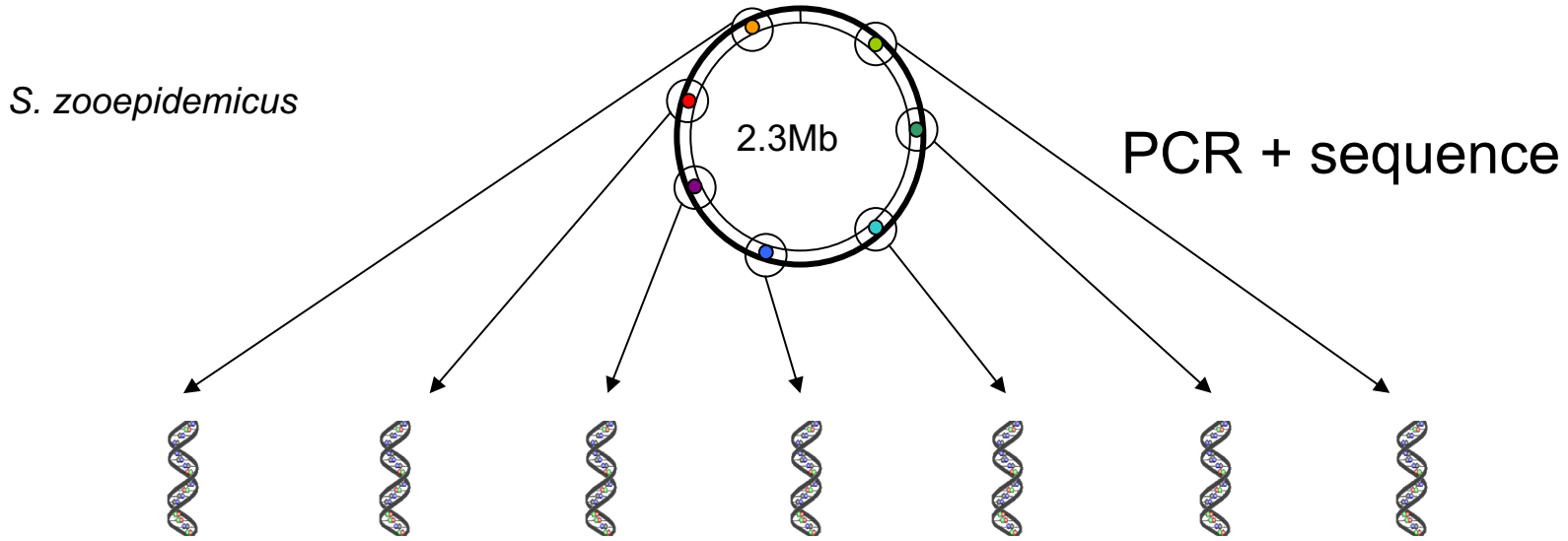



~500bp fragments

| ST | *arcC* | *nrdE* | *proS* | *spi* | *tdk* | *tpi* | *yqiL* |
|---|---|---|---|---|---|---|---|
| 6 | 2 | 4 | 4 | 6 | 4 | 5 | 6 |
| 80 | 2 | 4 | 4 | 7 | 4 | 5 | 20 |
| 133 | 2 | 4 | 4 | 7 | 5 | 5 | 20 |
| 137 | 16 | 4 | 4 | 7 | 4 | 3 | 20 |
| 149 | 16 | 4 | 4 | 7 | 4 | 5 | 20 |

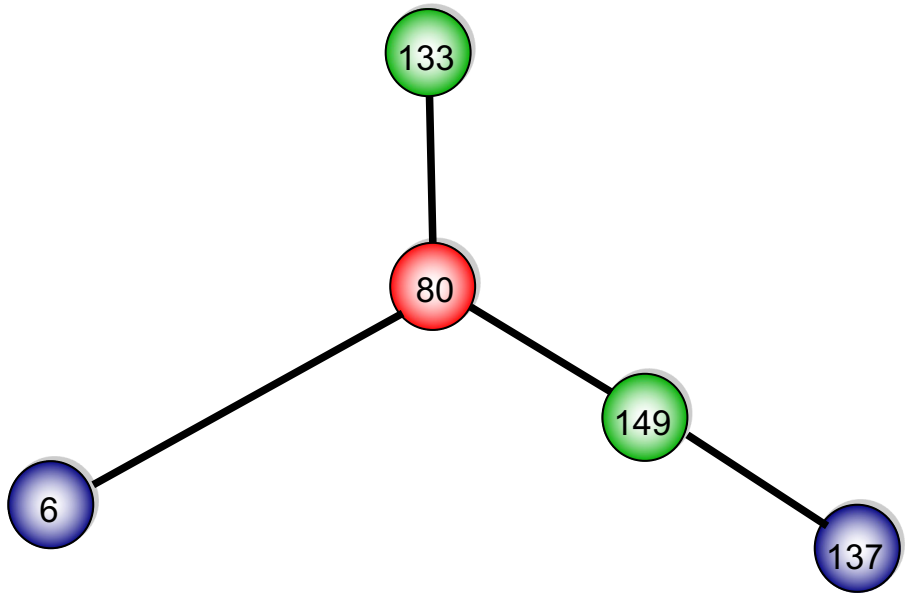
Summarises ~3500 bp DNA sequence into 7 integers (or 1 ST number)

# The principles of Multi-Locus Sequence Typing (MLST)




| ST | *arcC* | *nrdE* | *proS* | *spi* | *tdk* | *tpi* | *yqiL* |
|---|---|---|---|---|---|---|---|
| 6 | 2 | 4 | 4 | 6 | 4 | 5 | 6 |
| 80 | 2 | 4 | 4 | 7 | 4 | 5 | 20 |
| 133 | 2 | 4 | 4 | 7 | 5 | 5 | 20 |
| 137 | 16 | 4 | 4 | 7 | 4 | 3 | 20 |
| 149 | 16 | 4 | 4 | 7 | 4 | 5 | 20 |

# The principles of Multi-Locus Sequence Typing (MLST)

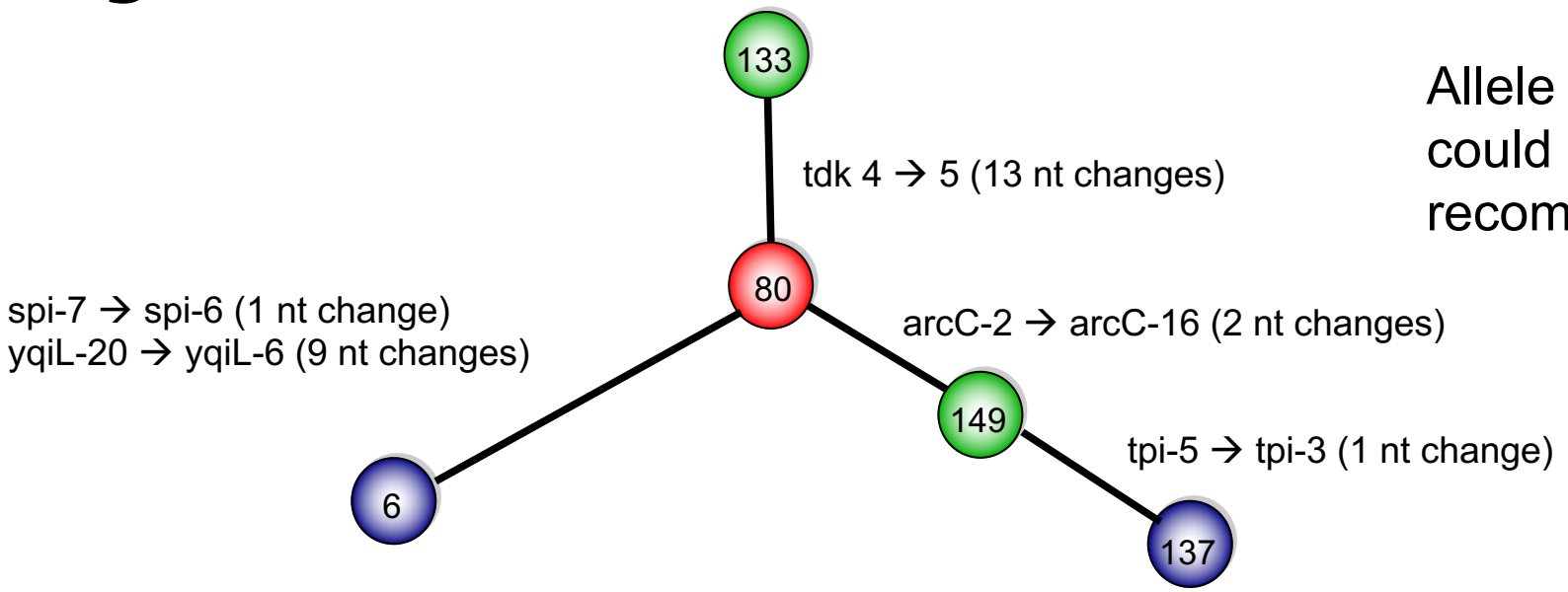



Related STs can be grouped together into clonal complexes, named after the central profile.

These often persist temporally and geographically.

| ST | *arcC* | *nrdE* | *proS* | *spi* | *tdk* | *tpi* | *yqiL* |
|---|---|---|---|---|---|---|---|
| 6 | 2 | 4 | 4 | 6 | 4 | 5 | 6 |
| 80 | 2 | 4 | 4 | 7 | 4 | 5 | 20 |
| 133 | 2 | 4 | 4 | 7 | 5 | 5 | 20 |
| 137 | 16 | 4 | 4 | 7 | 4 | 3 | 20 |
| 149 | 16 | 4 | 4 | 7 | 4 | 5 | 20 |

# Analysing alleles rather than sequences directly can mitigate the effects of recombination on tree topology



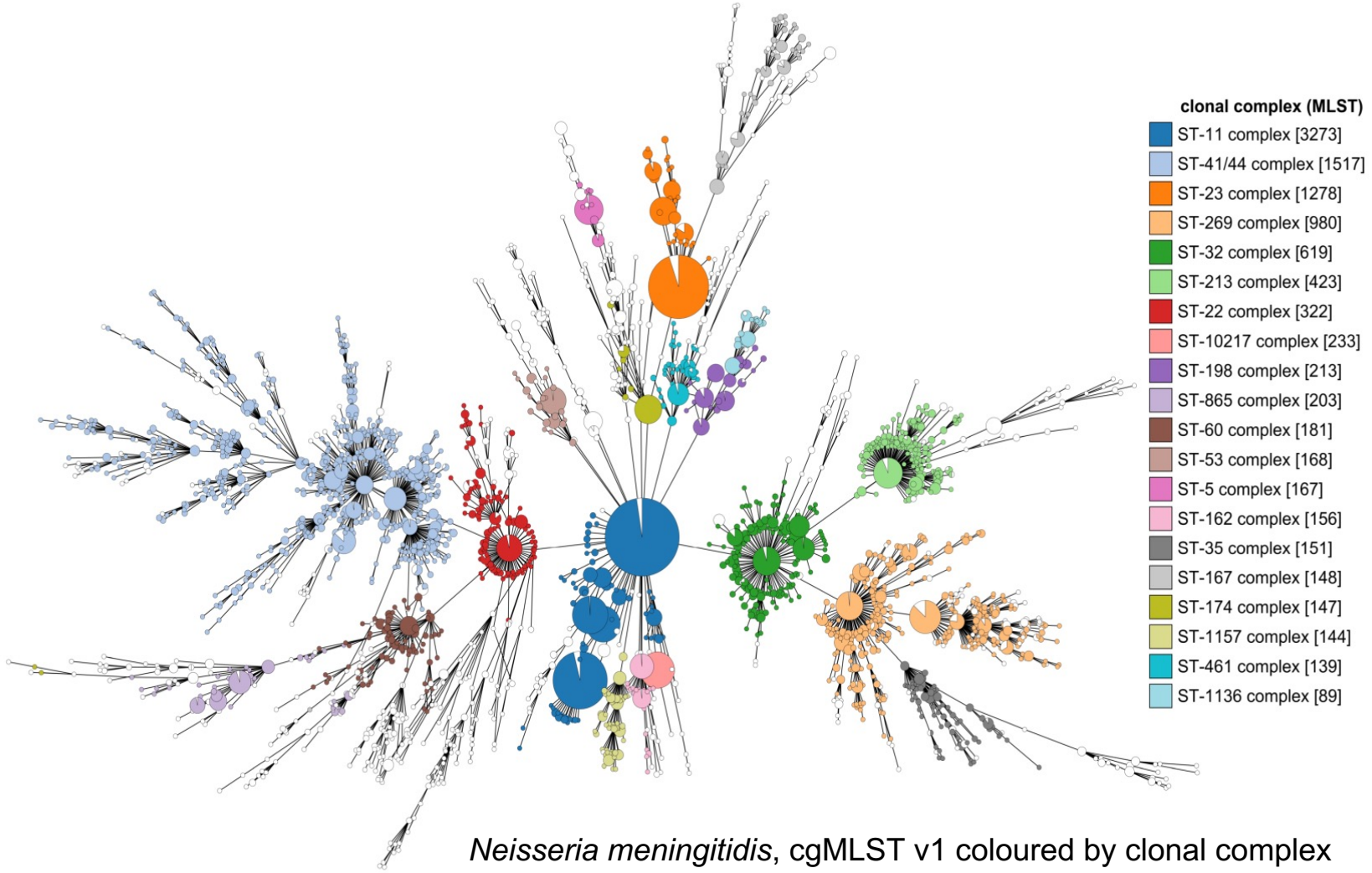

Allele change from tdk-4 to tdk-5 could be 13 mutation events or 1 recombination event

| ST | *arcC* | *nrdE* | *proS* | *spi* | *tdk* | *tpi* | *yqiL* |
|---|---|---|---|---|---|---|---|
| 6 | 2 | 4 | 4 | 6 | 4 | 5 | 6 |
| 80 | 2 | 4 | 4 | 7 | 4 | 5 | 20 |
| 133 | 2 | 4 | 4 | 7 | 5 | 5 | 20 |
| 137 | 16 | 4 | 4 | 7 | 4 | 3 | 20 |
| 149 | 16 | 4 | 4 | 7 | 4 | 5 | 20 |

# cgMLST facilitates high-resolution clustering by extending MLST to whole genome sequence data

The gene-by-gene approach

- More loci (hundreds or thousands) compared to usually 7 for MLST
- Complete coding sequences – including stop codon
- Missing data – most draft genomes will be missing some loci




*Neisseria meningitidis*, cgMLST v1 coloured by clonal complex

Bratcher *et al.* 2014 *BMC Genomics* 15:1138

# cgMLST scheme design principles

## Personal recommendation:

- Gene present in >98% isolates
- Allow up to 25 (or 2%) missing loci for profile definition

It is preferable to have fewer loci that are robustly identified than more loci that are sometimes missing.

# There is a problem with missing data when assigning cgSTs

Conventional 7-locus MLST is straightforward – most isolates have a complete allelic profile

| ST | abcZ | adk | aroE | fumC | gdh | pdhC | pgm |
|---|---|---|---|---|---|---|---|
| 11 | 2 | 3 | 4 | 3 | 8 | 4 | 6 |

cgMLST profiles usually have missing data (~1600 loci from draft genomes)

| NEIS0001 | NEIS0004 | NEIS0005 | NEIS0006 | NEIS0007 | NEIS0008 | NEIS0009 | NEIS0010 | NEIS0011 | NEIS0012 | NEIS0013 | NEIS0014 | NEIS0015 | NEIS0016 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 8 | 18 | 10 | 64 | | 10 | 13 | 12 | 13 | 9 | 2 | 13 | 5 |

# Defining cgSTs – the problem with missing data

cgST profiles can be assigned with a ‘N’ for missing loci

| NEIS0001 | NEIS0004 | NEIS0005 | NEIS0006 | NEIS0007 | NEIS0008 | NEIS0009 | NEIS0010 | NEIS0011 | NEIS0012 | NEIS0013 | NEIS0014 | NEIS0015 | NEIS0016 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 8 | 18 | 10 | 64 | N | 10 | 13 | 12 | 13 | 9 | 2 | 13 | 5 |
| 2 | 8 | 18 | 10 | 64 | 14 | 10 | 13 | 12 | 13 | 9 | 2 | 13 | 5 |

- ‘N’ matches any allele
- Missing loci must be limited
  - profiles can appear artificially related
- PubMLST defines cgST for profiles with <= 25 missing loci

# Records with missing data can appear more related than they are

| NEIS0001 | NEIS0004 | NEIS0005 | NEIS0006 | NEIS0007 | NEIS0008 | NEIS0009 | NEIS0010 | NEIS0011 | NEIS0012 | NEIS0013 | NEIS0014 | NEIS0015 | NEIS0016 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 8 | 18 | 10 | 64 | | 10 | 13 | 12 | 13 | 9 | 2 | 13 | 5 |
| 2 | 12 | 18 | 10 | 64 | 3 | 10 | 4 | | 13 | 2 | 2 | 13 | 5 |

This is counted as 3 differences, but it could be up to 5!

# PubMLST/BIGSdb uses BLAST for allele calling

## Strategy 1: Naïve


- BLAST genome sequence against database of all alleles
- One locus at a time
- BLAST database constructed on-the-fly

| Locus | Allele |
|---|---|
| LOCUS0001 | 5 |
| LOCUS0002 | |
| LOCUS0003 | |
| LOCUS0004 | |

Jolley & Maiden (2010) *BMC Bioinformatics.* 11:595

# Allele calling strategy 1: Naïve

Searching against all alleles, locus-by-locus

- Sensitive
  - Will identify every known allele
- Easy to program

- Multiple BLAST queries needed
- Locus definition can drift
- Gets slower as more alleles added

SLOW!!

# Allele calling strategy 2: Type alleles

BLAST against one example allele

- Constrains search space
  - Ensures locus definition doesn't drift
  - Good for cgMLST and other routine typing
- Reduced search space means all loci can be searched together

- Only alleles within a defined distance of the type allele will be found
- Not good if there are multiple allele classes for a locus

FAST!!!

# Allele calling strategy 3: Exemplar alleles

BLAST against multiple example alleles covering known diversity
Exemplars on PubMLST defined such that all known alleles are within 3% identity of an exemplar allele of the same length.

- Constrains search space
- Identifies all known alleles – even if highly variable
- Reduced search space means all loci can be searched together

- Locus definition can drift

FAST!!!

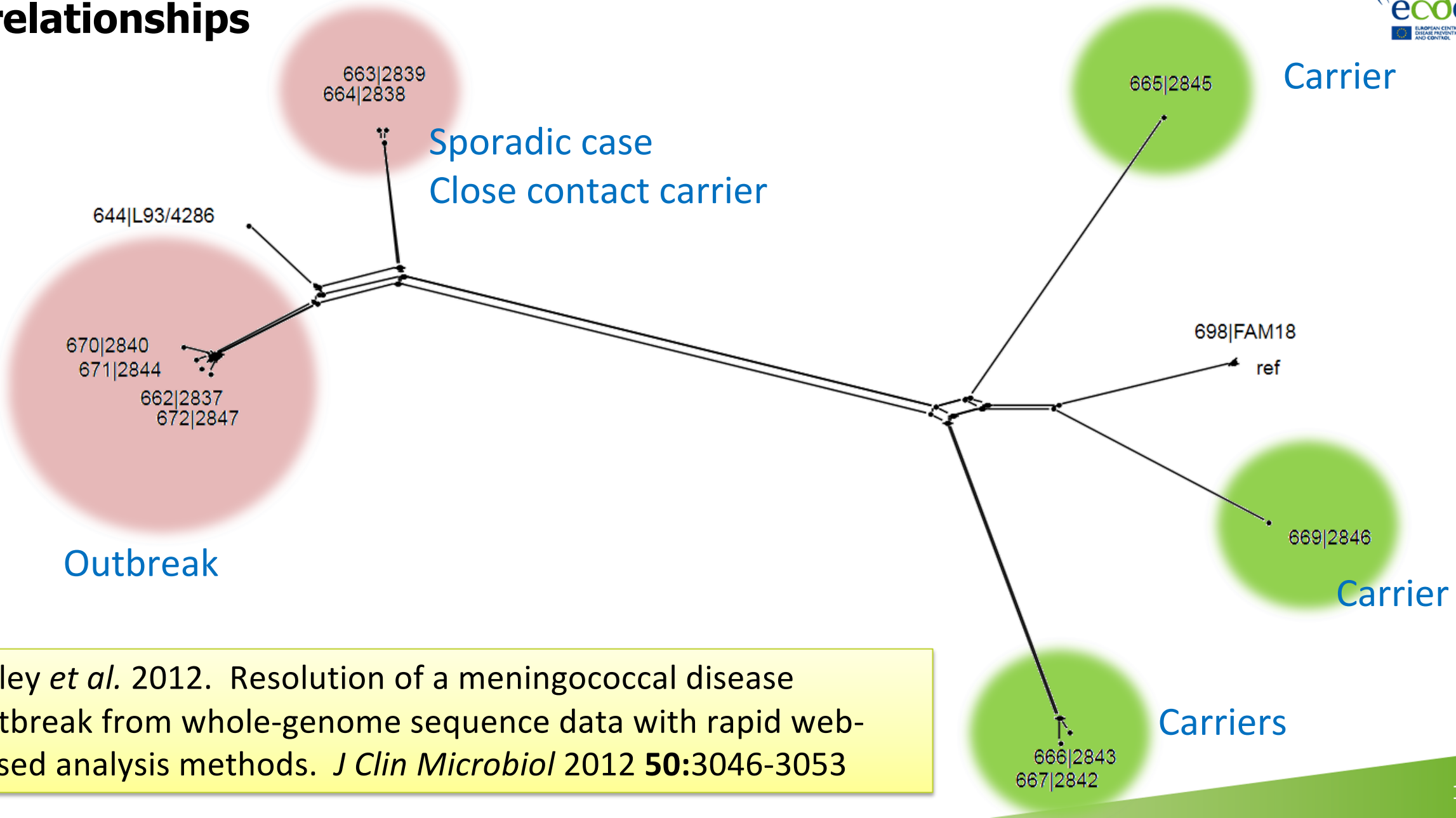
# Genomes can be rapidly compared using gene-by-gene variant numbering approach

| 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 |
| NMC0031 | conserved hypothetical protein | 1209 | 31875 | 1 | 2 | 2 | 1 | 2 | 2 | I | 2 | 2 | 2 | 1 | 4 | 5 |

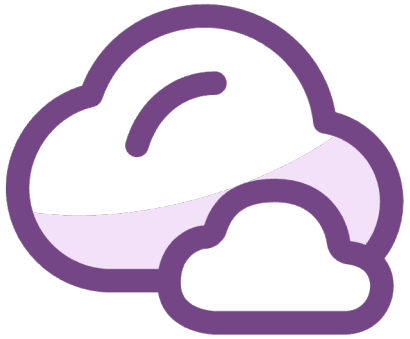
# Distance matrix shows number of genes that are different between every pair of isolates

```
#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;
```

# NeighborNet can visualise 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



# What are the advantages of gene-by-gene approaches like MLST for genomic surveillance?

# Strengths and limitations of gene-by-gene approaches to genomic surveillance

## Strengths

- Standardisation and comparability
- Scalability – more loci: higher resolution
- Computationally efficient
- Intuitive
- Robust to moderate recombination
- Database-driven integration

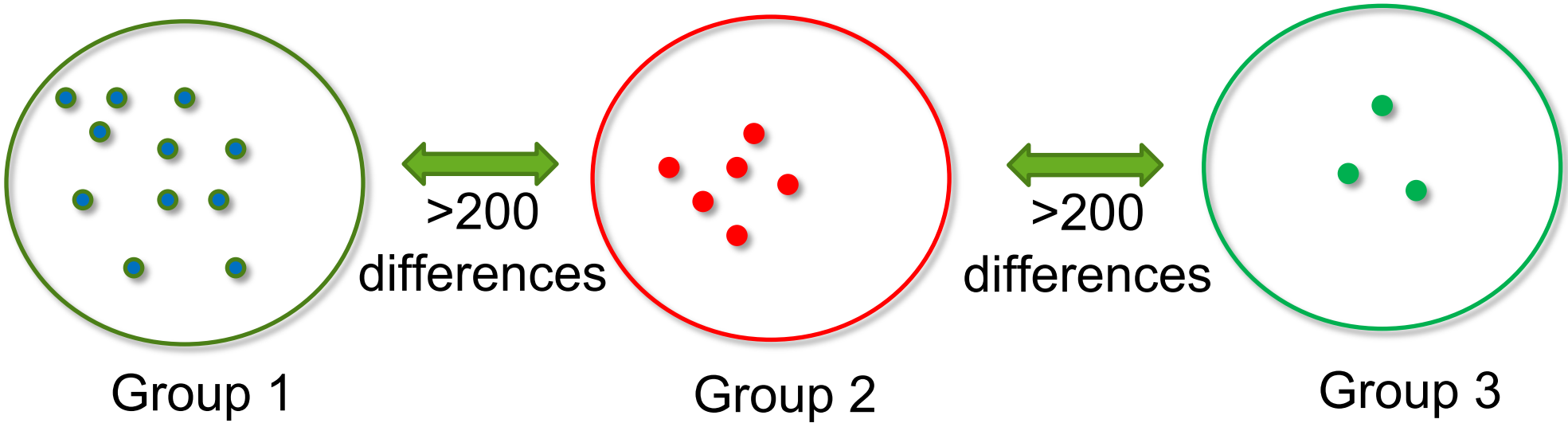
## Limitations

- Requires centrally curated database required
- Resolution for applications requiring very high discrimination – only core loci used
- Relies on good genome assemblies
- Does not capture nucleotide-level variation directly – less suitable for monomorphic species

# Classification schemes can be used to cluster profiles

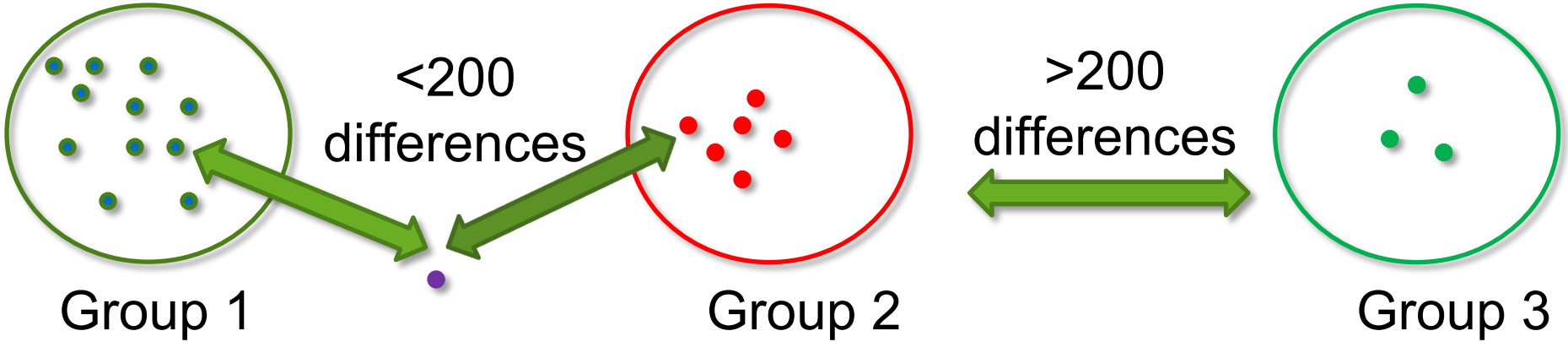
Group cgSTs based on different thresholds (200, 100, 50, 25 loci)

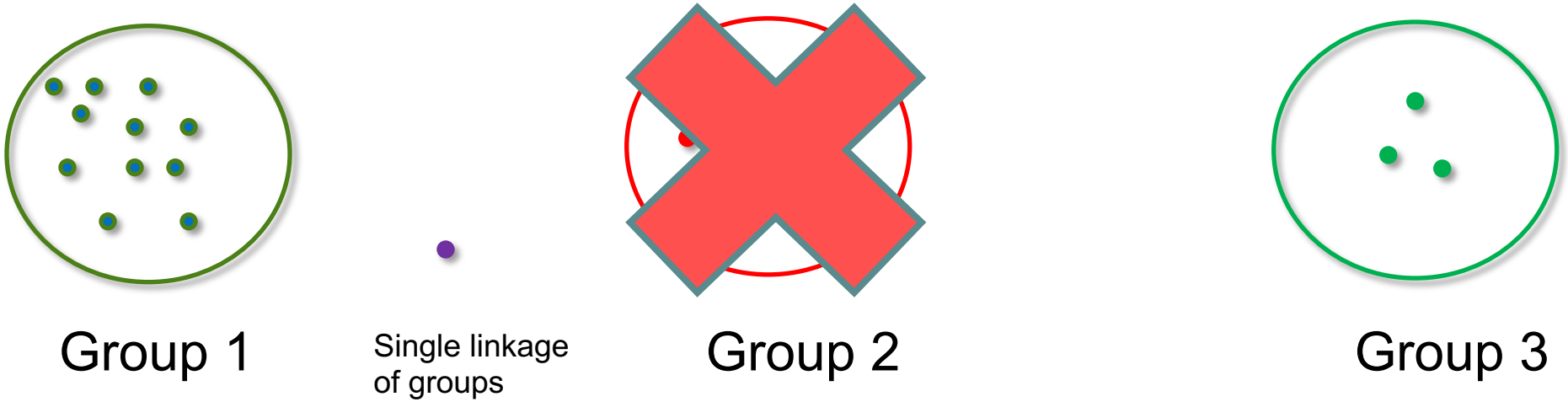
Nm_cgc_200 (*Neisseria meningitidis* core genome cluster 200 threshold)



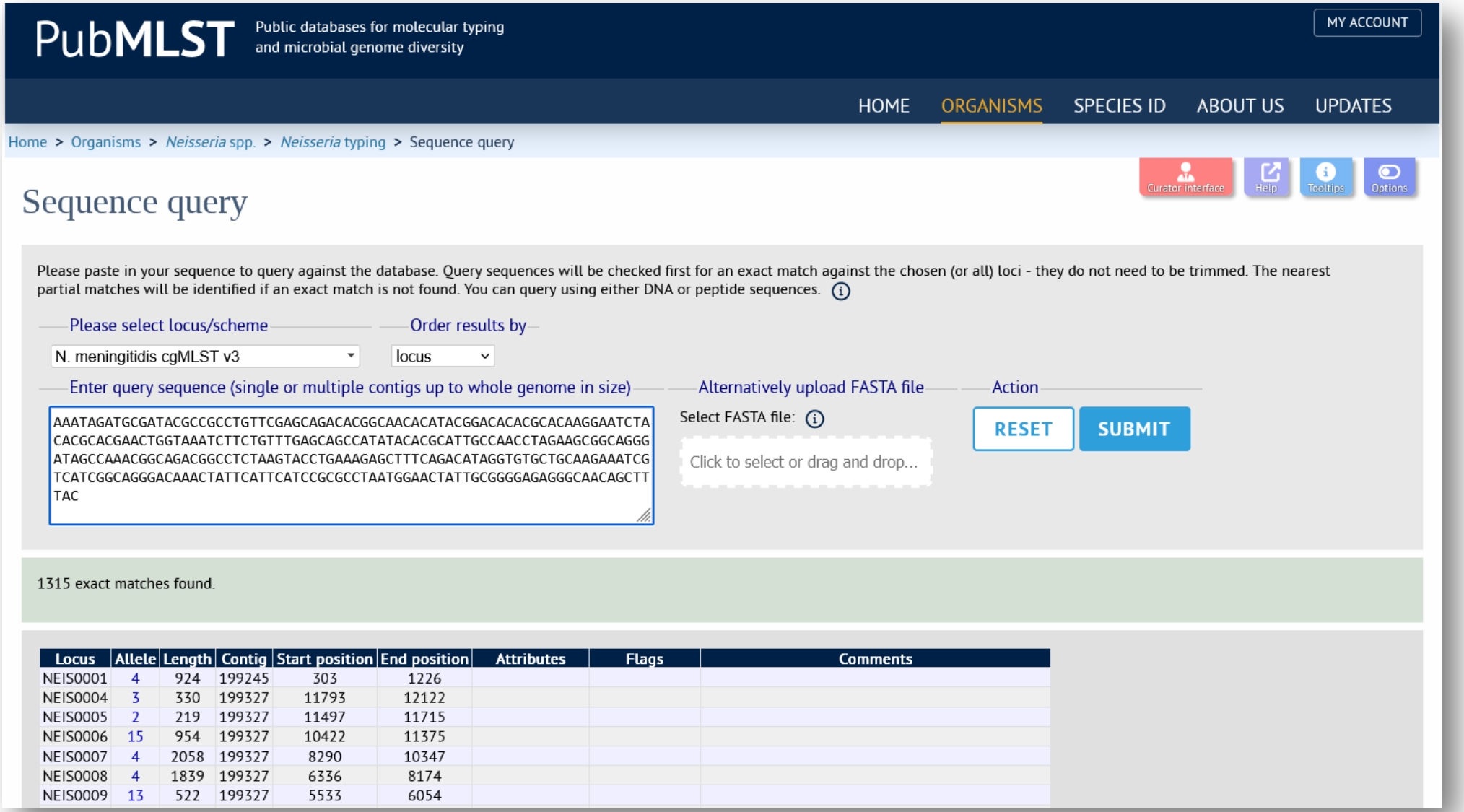

Single linkage clusters

# Groups merge when a new record is within the threshold distance of 2 or more groups



# cgMLST clusters can be used to identify nearest neighbours



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

Curator interface Help Tooltips Options

## 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: N. meningitidis cgMLST v3

Order results by: locus

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

```
AAATAGATGCGATACGCCGCCTGTTCGAGCAGACACGGCAACACATACGGACACACGCACAAGGAATCTA
CACGCACGAACTGGTAAATCTTCTGTTTGAGCAGCCATATACACGCATTGCCAACCTAGAAGCGGCAGGG
ATAGCCAAACGGCAGACGGCCTCTAAGTACCTGAAAGAGCTTTCAGACATAGGTGTGCTGCAAGAAATCG
TCATCGGCAGGGACAAACTATTCATTCATCCGCGCCTAATGGAACTATTGCGGGGAGAGGGCAACAGCTT
TAC
```

Alternatively upload FASTA file

Select FASTA file:

Click to select or drag and drop...

Action: RESET SUBMIT

1315 exact matches found.

| Locus | Allele | Length | Contig | Start position | End position | Attributes | Flags | Comments |
|---|---|---|---|---|---|---|---|---|
| NEIS0001 | 4 | 924 | 199245 | 303 | 1226 | | | |
| NEIS0004 | 3 | 330 | 199327 | 11793 | 12122 | | | |
| NEIS0005 | 2 | 219 | 199327 | 11497 | 11715 | | | |
| NEIS0006 | 15 | 954 | 199327 | 10422 | 11375 | | | |
| NEIS0007 | 4 | 2058 | 199327 | 8290 | 10347 | | | |
| NEIS0008 | 4 | 1839 | 199327 | 6336 | 8174 | | | |
| NEIS0009 | 13 | 522 | 199327 | 5533 | 6054 | | | |

# cgMLST clusters can be used to identify nearest neighbours

| | | | | | |
|---|---|---|---|---|---|
| NEIS0010 | 4 | 1326 | 199327 | 3792 | 5117 |
| NEIS0011 | 16 | 840 | 199327 | 2742 | 3581 |
| NEIS0012 | 4 | 1167 | 199327 | 1579 | 2745 |

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.

Text Excel

## N. meningitidis cgMLST v3

### Matching profiles

**Closest profile:** cgST-56
**Mismatches:** 1
**Loci matched:** 1328/1329 (99.9%)

### Similar profiles (determined by classification schemes)

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

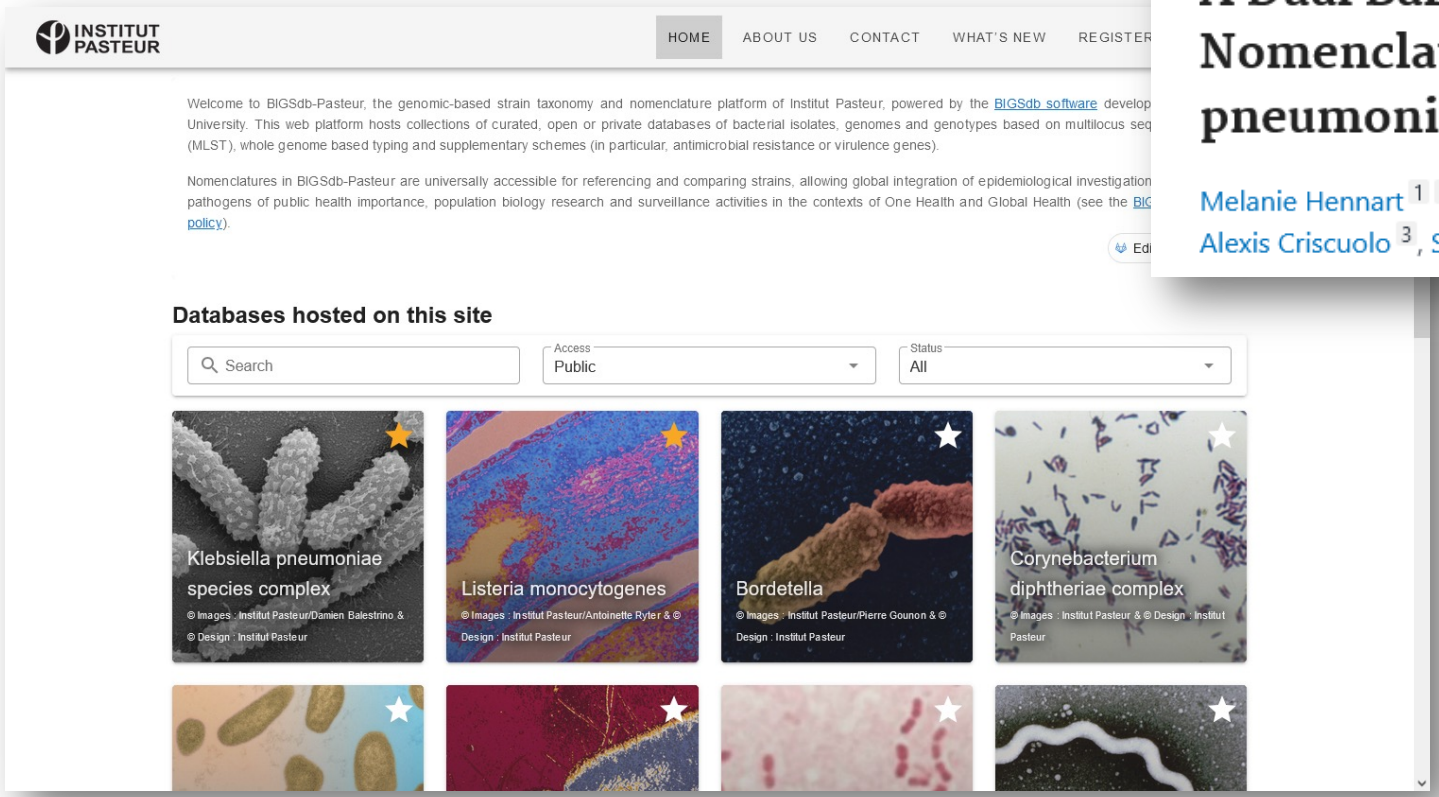
cgST-56 belongs to the following clusters:

| Classification scheme | Clustering method | Mismatch threshold | Status | Group | Profiles | Isolates |
|---|---|---|---|---|---|---|
| Nm_cgc3_200 | Single-linkage | 200 | experimental | 8 | 3775 | PubMLST isolates 2346 |
| Nm_cgc3_100 | Single-linkage | 100 | experimental | 20 | 2225 | PubMLST isolates 1370 |
| Nm_cgc3_50 | Single-linkage | 50 | experimental | 726 | 164 | PubMLST isolates 101 |
| Nm_cgc3_25 | Single-linkage | 25 | experimental | 56 | 2 | PubMLST isolates 2 |
| Nm_cgc3_10 | Single-linkage | 10 | experimental | 56 | 2 | PubMLST isolates 2 |

# Single-linkage clustering is not stable over time

- Consistent at any point in time – can be useful for identifying outbreaks.
- Thresholds represent minimum distance between groups.
- Individual pairs within groups may be further apart than the threshold distance.
- Groups merge and get larger over time.

# Life Identification Numbers (LIN) using cgMLST provide a stable nomenclature

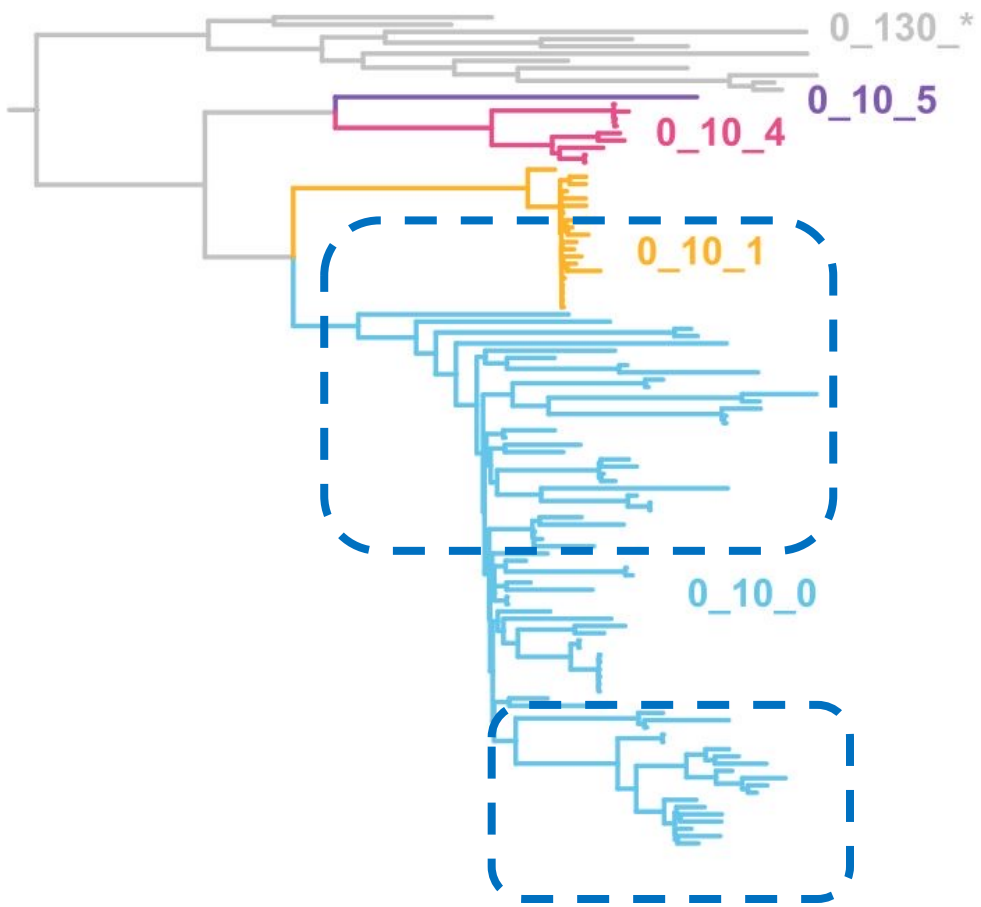


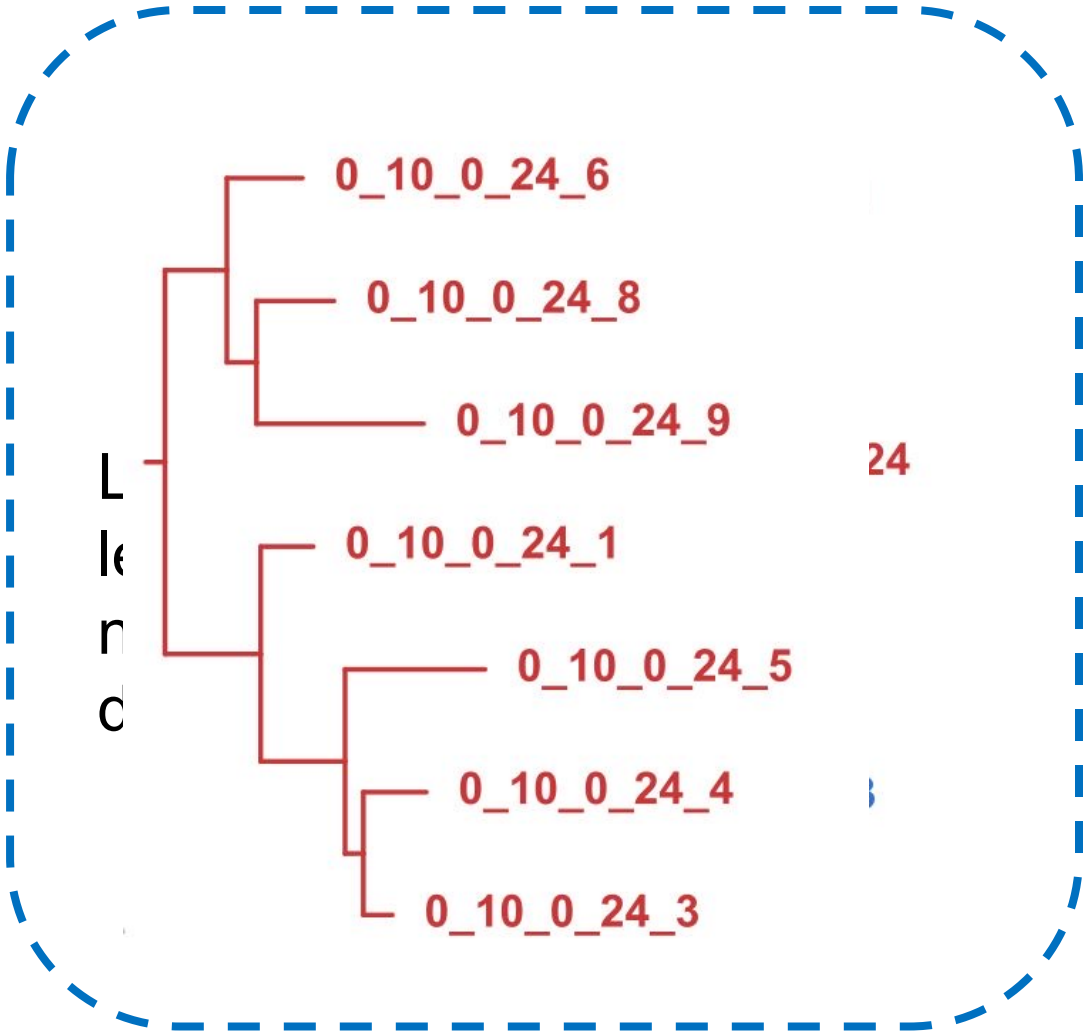

Mol Biol Evol. 2022 Jul 2;39(7):msac135. doi: 10.1093/molbev/msac135.

## A Dual Barcoding Approach to Bacterial Strain Nomenclature: Genomic Taxonomy of Klebsiella pneumoniae Strains

Melanie Hennart [1 2], Julien Guglielmini [3], Sébastien Bridel [1], Martin C J Maiden [4], Keith A Jolley [4], Alexis Criscuolo [3], Sylvain Brisse [1]

# LIN codes provide a stable nomenclature to cgMLST clusters




van Rensburg *et al.* 2024 *Micro Genom* 10(8):001280

# We can now use LIN codes to identify closest matching genomes in PubMLST following a sequence query

## cgMLST v1

### Matching profiles

**Closest profile:** cgST-1448
**Mismatches:** 164
**Loci matched:** 873/1037 (84.2%)

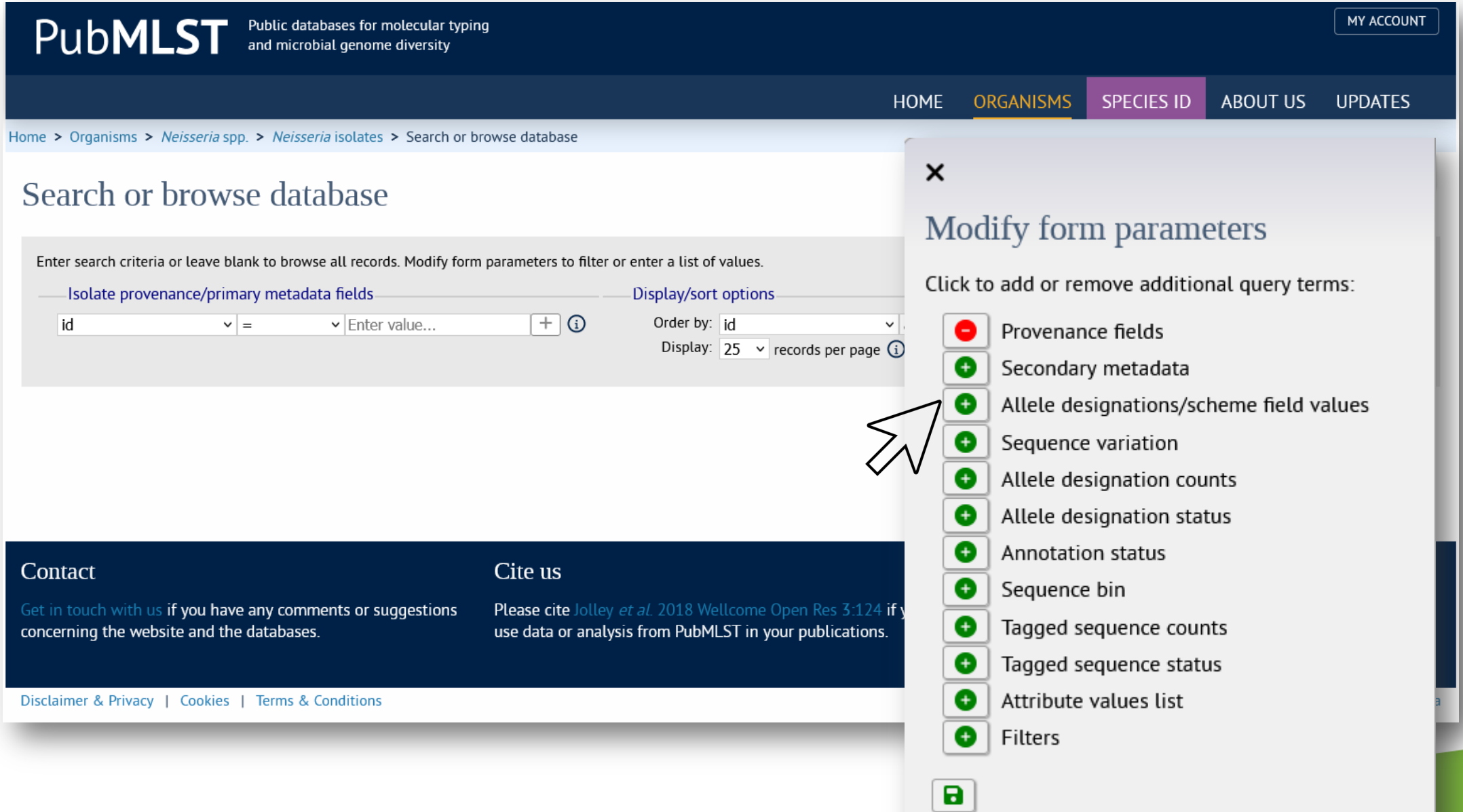
### Similar profiles (determined by LIN codes)

**Partial LIN code:** 0_1_0_2_0_0

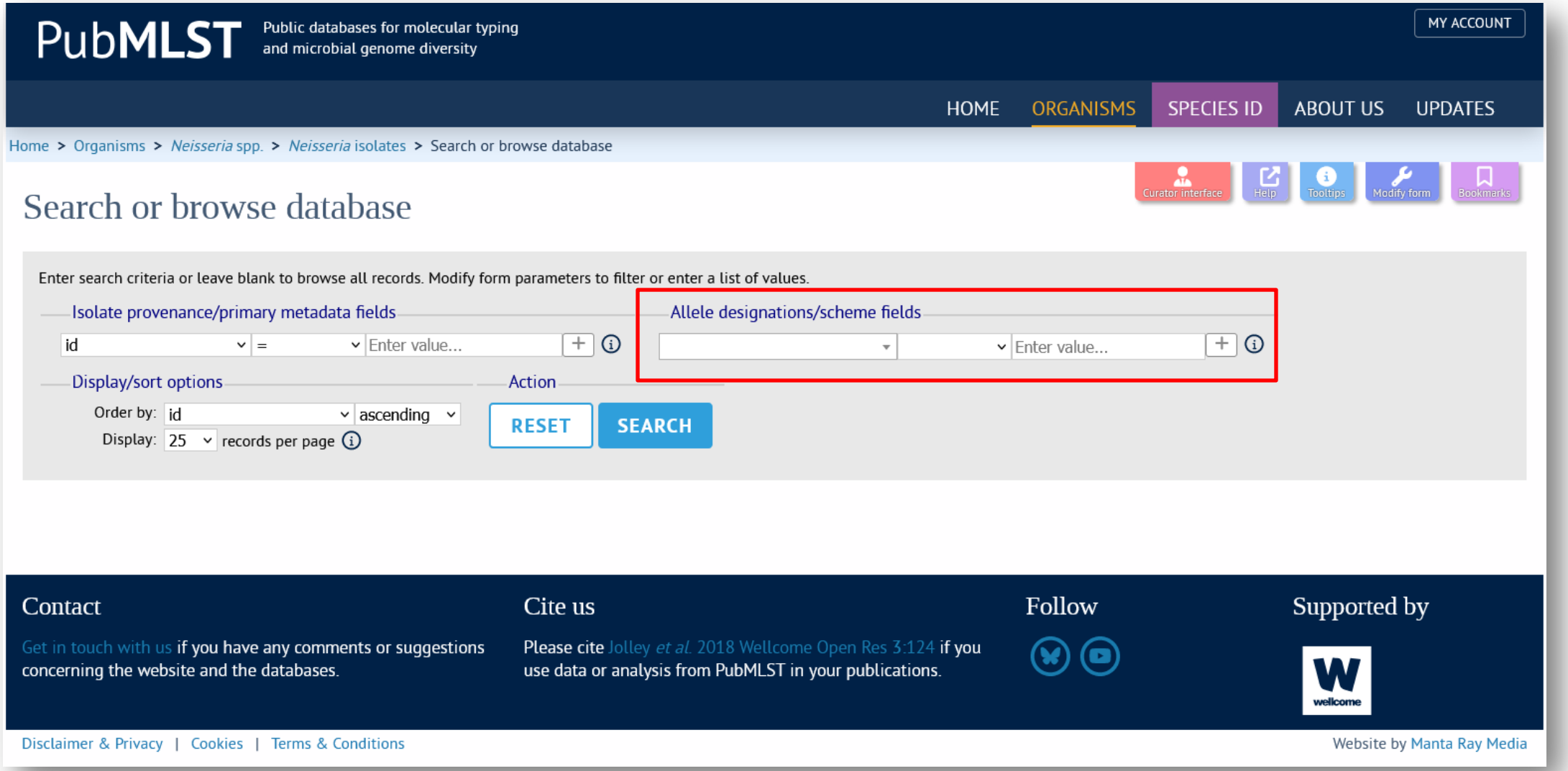
Show all thresholds

| Prefix | Threshold | Matching profiles | Isolates |
|---|---|---|---|
| 0_1 | 800 | 1960 | PubMLST isolates 663 |
| 0_1_0 | 630 | 1153 | PubMLST isolates 406 |
| 0_1_0_2 | 480 | 414 | PubMLST isolates 141 |
| 0_1_0_2_0 | 280 | 62 | PubMLST isolates 25 |
| 0_1_0_2_0_0 | 200 | 62 | PubMLST isolates 25 |

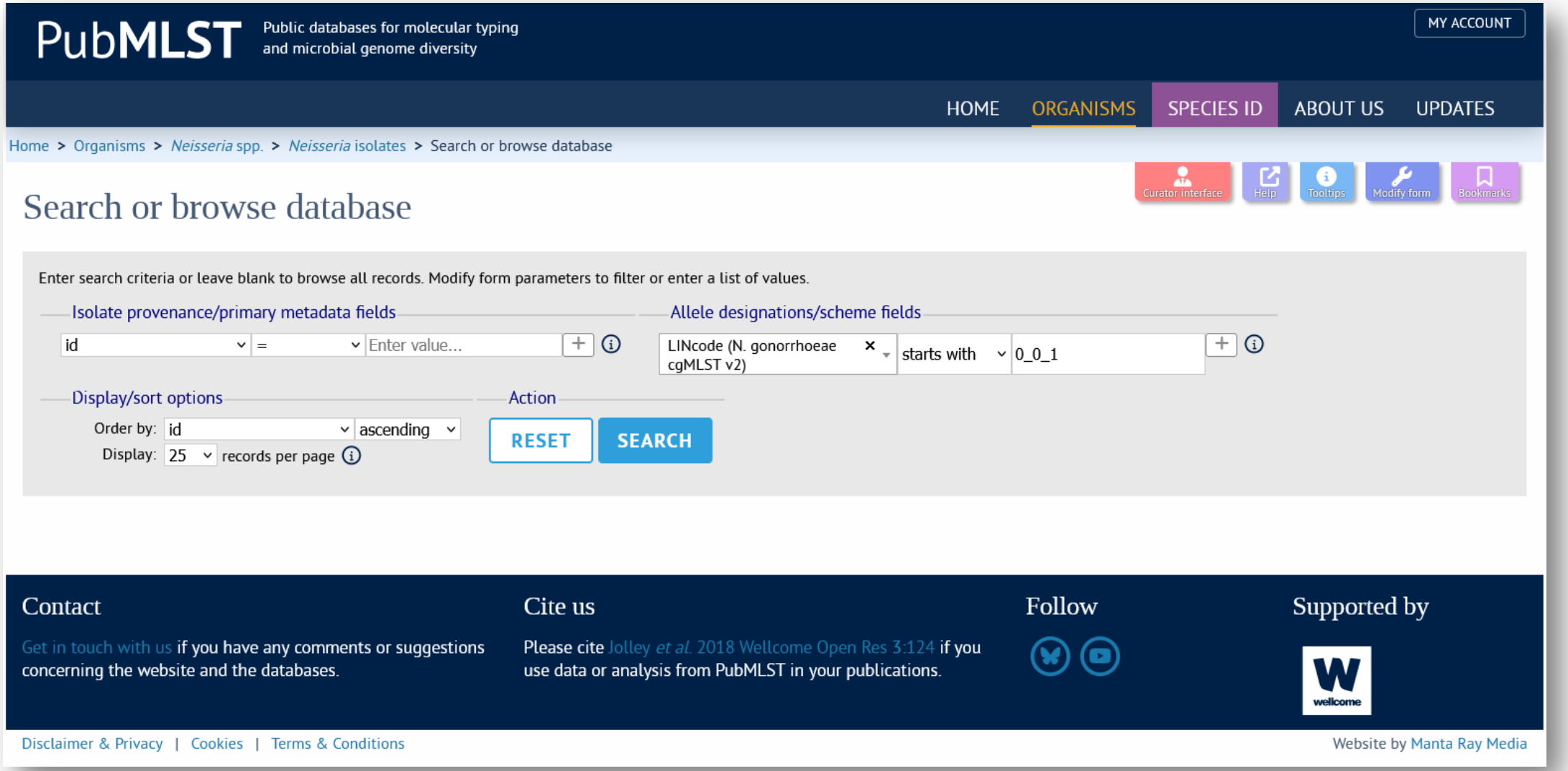
# Search by LIN code prefix or nickname



# Search by LIN code prefix or nickname



# Search by LIN code prefix or nickname



# Search by LIN code prefix or nickname


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

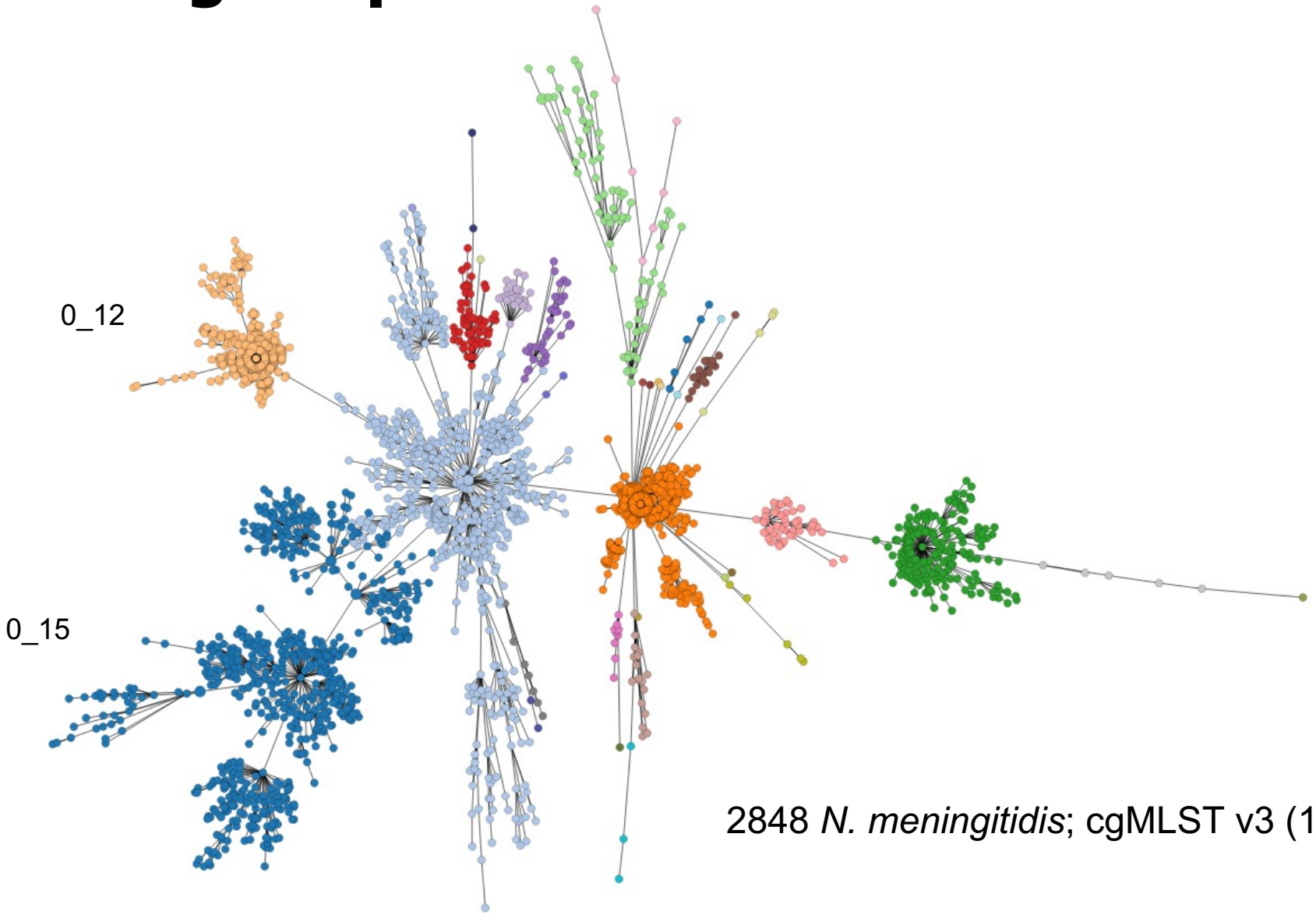
Your projects | Bookmark query

Add to project | Bookmark

« ‹ 1 2 3 4 5 › »

| Isolate fields | | | | | | | | | | | | Seqbin size (bp) | Contigs | MLST | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| id | isolate | aliases | country | continent | region | town or city | year | date sampled | disease | species | capsule group | | | abcZ | adk | aroE | fumC | gdh | pdhC | pgm | ST | clonal complex | PorA |
| 31509 | POL_11_148 | 10900_8_19 | Poland | Europe | | | 2011 | | | Neisseria gonorrhoeae | | 2,200,098 | 185 | 126 | 39 | 67 | 238 | 146 | 153 | 133 | 10932 | | 18 |
| 31635 | POL_10_5 | 10868_8_75 | Poland | Europe | | | 2010 | | | Neisseria gonorrhoeae | | 2,197,160 | 203 | 126 | 39 | 67 | 238 | 146 | 153 | 133 | 10932 | | 18 |
| 31657 | POL_12_159 | 10900_8_32; ERR352724 | Poland | Europe | | | 2012 | | | Neisseria gonorrhoeae | | 2,198,827 | 182 | 126 | 39 | 67 | 238 | 146 | 153 | 133 | 10932 | | 18 |
| 34773 | ATL_2011_01_17 | | USA | North America | Georgia | Atlanta | 2011 | | | Neisseria gonorrhoeae | | 2,145,634 | 353 | 126 | 39 | 67 | 238 | 146 | 153 | 133 | 10932 | | 18 |
| 37107 | GCGS0376 | 12CFX_C_015; 3416STDY6016787 | USA | North America | Missouri | | 2012 | | | Neisseria gonorrhoeae | | 2,191,447 | 204 | 126 | 39 | 67 | 238 | 146 | 153 | 133 | 10932 | | 18 |
| 37152 | GCGS0428 | 11CFX_C_024; 3416STDY6016833 | USA | North America | Ohio | | 2011 | | | Neisseria gonorrhoeae | | 2,202,553 | 178 | 126 | 39 | 67 | 238 | 146 | 153 | 133 | 10932 | | 18 |
| 37187 | GCGS0429 | 12CFX_C_027; 3416STDY6016868 | USA | North America | Ohio | | 2012 | | | Neisseria gonorrhoeae | | 2,195,737 | 239 | 126 | 39 | 67 | 238 | 146 | 153 | 133 | 10932 | | 18 |
| 37292 | GCGS0363 | 11AZI_T_002; 3416STDY6016975 | USA | North America | Michigan | | 2011 | | | Neisseria gonorrhoeae | | 2,199,688 | 232 | 126 | 39 | 67 | 238 | 146 | 153 | 133 | 10932 | | 18 |

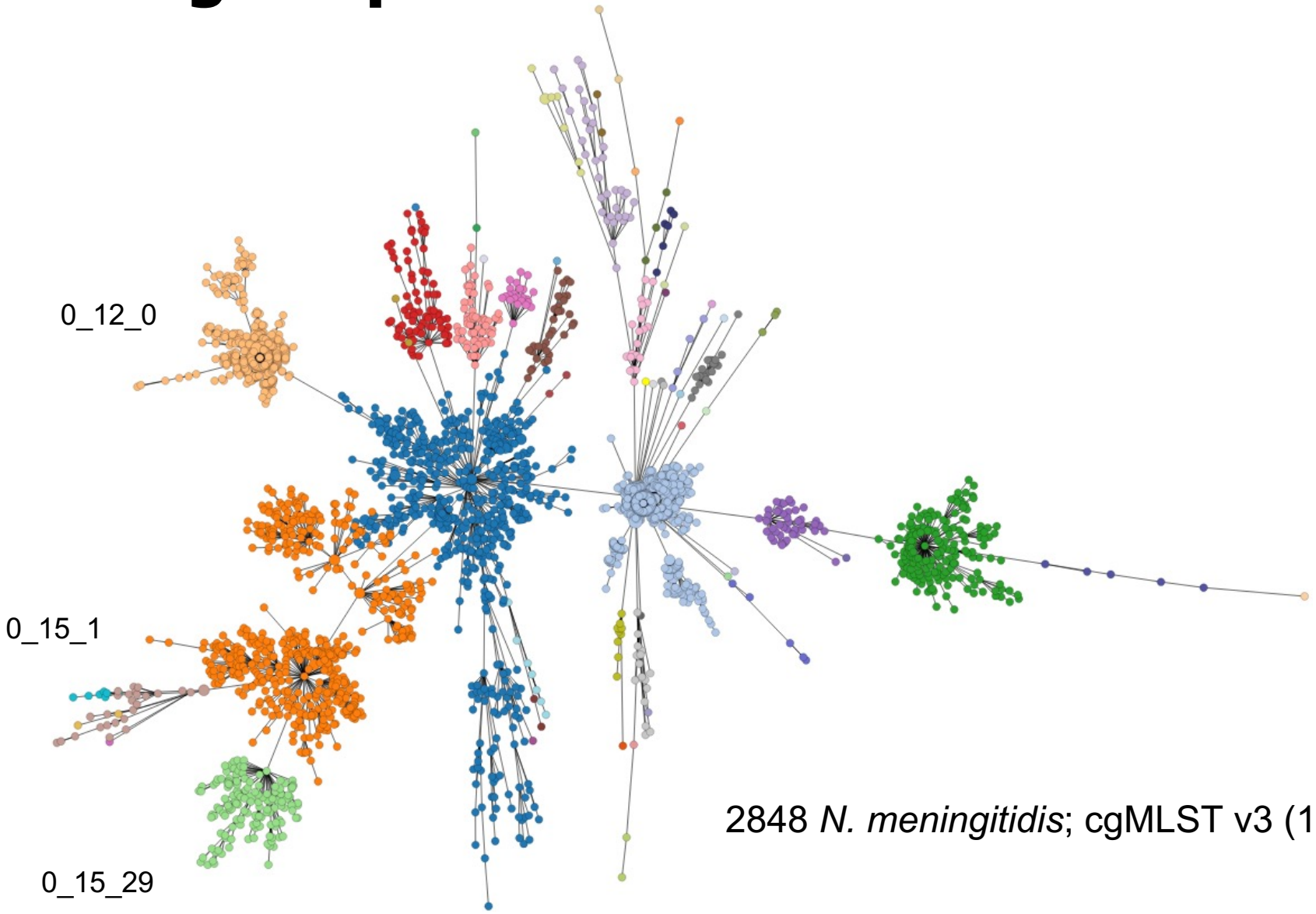
# LIN code prefixes can be overlaid on a minimum-spanning GrapeTree within PubMLST




LIN code [2]

2848 *N. meningitidis*; cgMLST v3 (1329 loci)

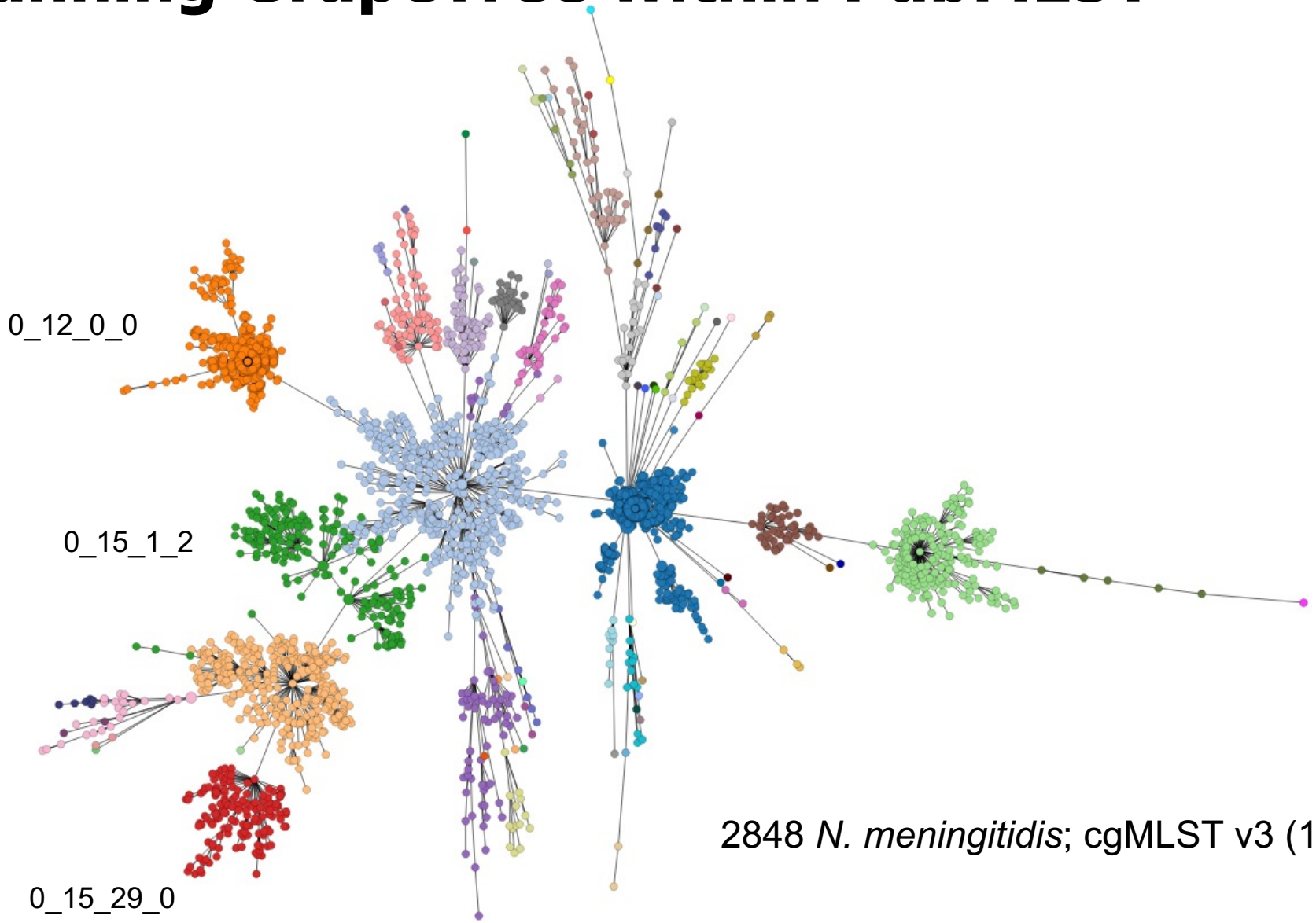
# LIN code prefixes can be overlaid on a minimum-spanning GrapeTree within PubMLST




LIN code [3]

2848 *N. meningitidis*; cgMLST v3 (1329 loci)

# LIN code prefixes can be overlaid on a minimum-spanning GrapeTree within PubMLST




LIN code [4]

2848 *N. meningitidis*; cgMLST v3 (1329 loci)

# Further reading


- Maiden *et al.* (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 11(10), 728-736.
- Jolley *et al.* (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3:124.
- Hennart *et al.* (2022). A dual barcoding approach to bacterial strain nomenclature: Genomic taxonomy of *Klebsiella pneumoniae* strains. *Mol Biol Evol* 39(7):msac135.

# Acknowledgements

**The creation of this training material was commissioned by ECDC to University of Oxford with the direct involvement of Keith Jolley**