Virtual Training 15: Focus on the Agents of Diphtheria

Practical exercices: using the BIGSdb-Pasteur platform for Corynebacterium diphtheriae genomics and molecular epidemiology - answers

Pr Sylvain Brisse, Institut Pasteur

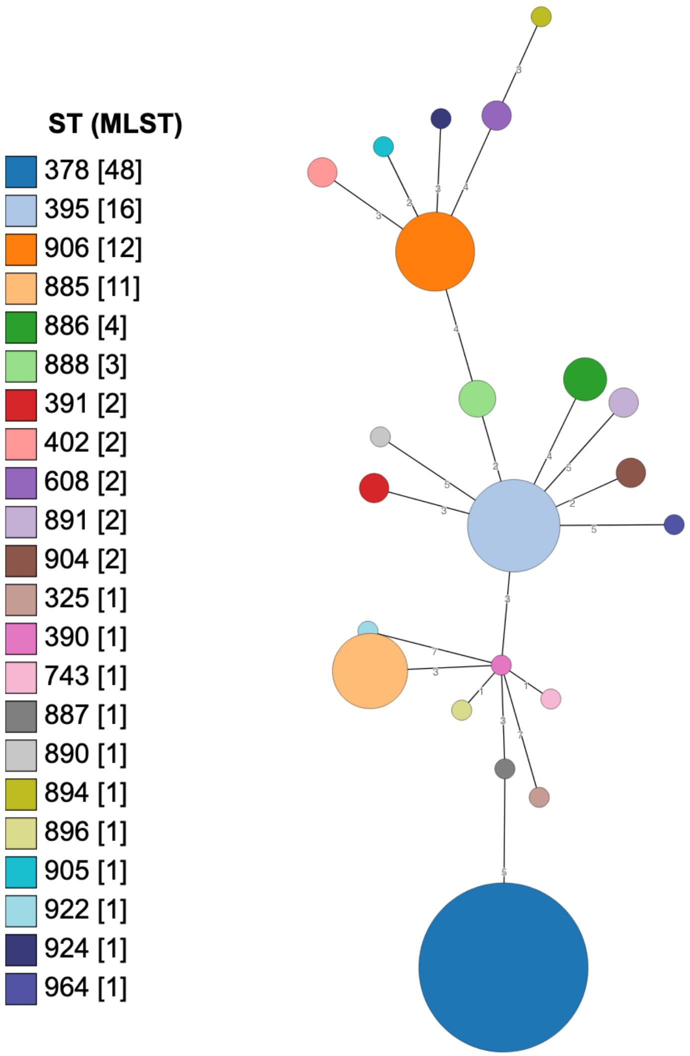
Practical exercices

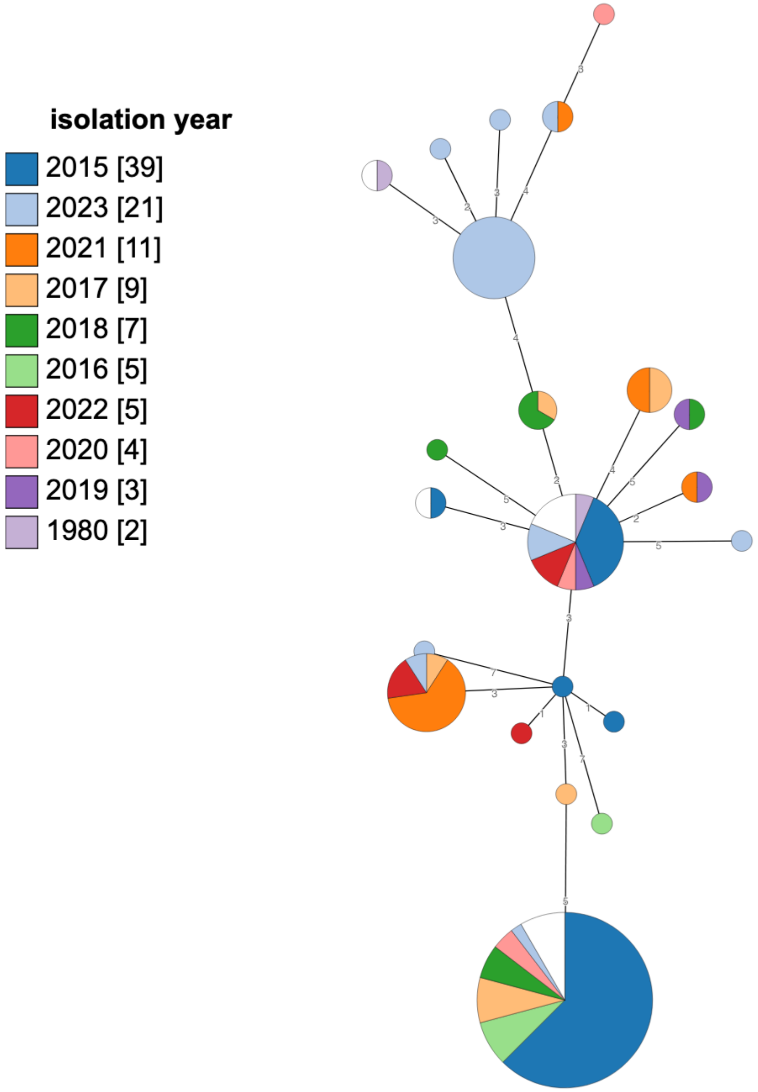
- **Please see the five proposed exercices and perform the analyses to answer the questions**
- **Please interpret the results in terms of diphtheria epidemiology and evolution**

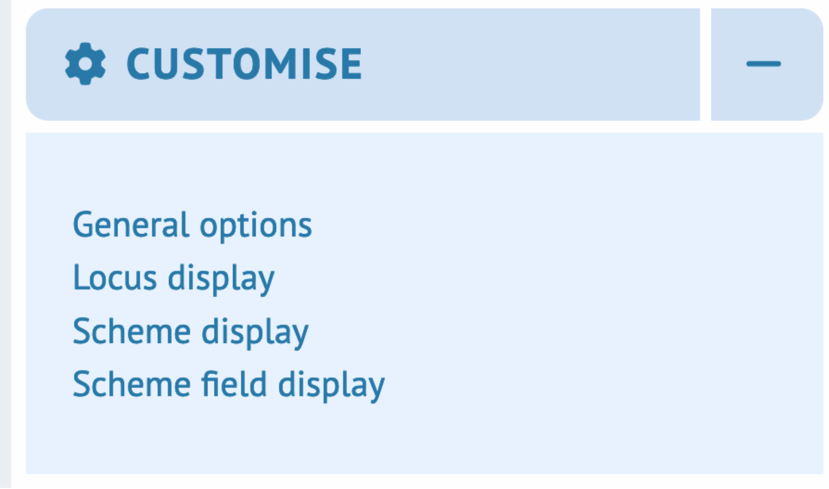
CdSC (*Corynebacterium diphtheriae* species complex) BIGSdb-Pasteur platform

- Search for public isolates from South Africa
- How are they distributed by MLST and year?
- Can we suspect several outbreaks occurred separately?
- How to find the evidence?

South Africa isolates, GrapeTree based on MLST (7 genes)

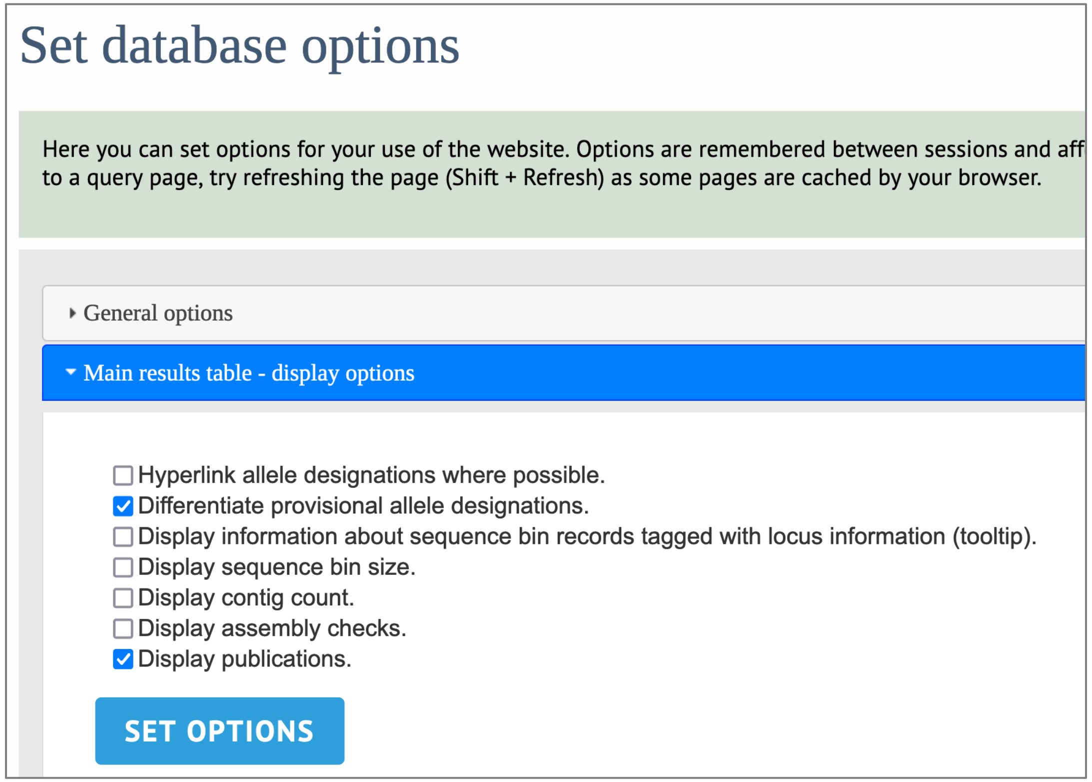






CUSTOMISE

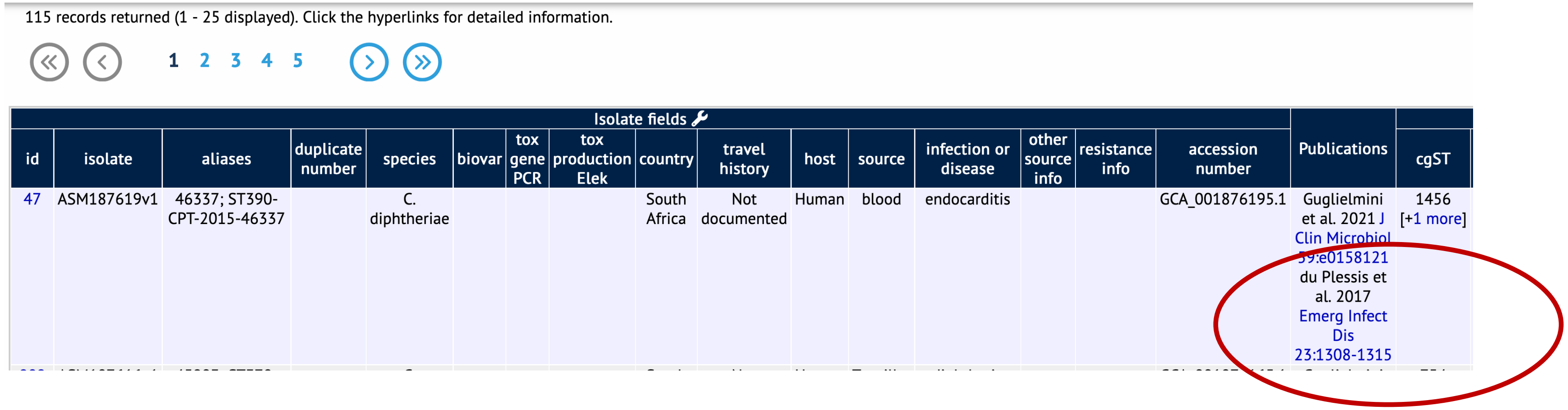
General options
Locus display
Scheme display
Scheme field display



Set database options

Here you can set options for your use of the website. Options are remembered between sessions and aff
to a query page, try refreshing the page (Shift + Refresh) as some pages are cached by your browser.

☐ 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 assembly checks.
☑ Display publications.



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

Isolate fields																Publications	cgST
id	isolate	aliases	duplicate number	species	biovar	tox gene PCR	tox production Elek	country	travel history	host	source	infection or disease	other source info	resistance info	accession number		
47	ASM187619v1	46337; ST390-CPT-2015-46337		C. diphtheriae				South Africa	Not documented	Human	blood	endocarditis			GCA_001876195.1	Guglielmini et al. 2021 J Clin Microbiol 59:e0158121 du Plessis et al. 2017 Emerg Infect Dis 23:1308-1315	1456 [+1 more]

Review > Emerg Infect Dis. 2025 Mar;31(3):417-426. doi: 10.3201/eid3103.241211.

Corynebacterium diphtheriae Infections, South Africa, 2015–2023

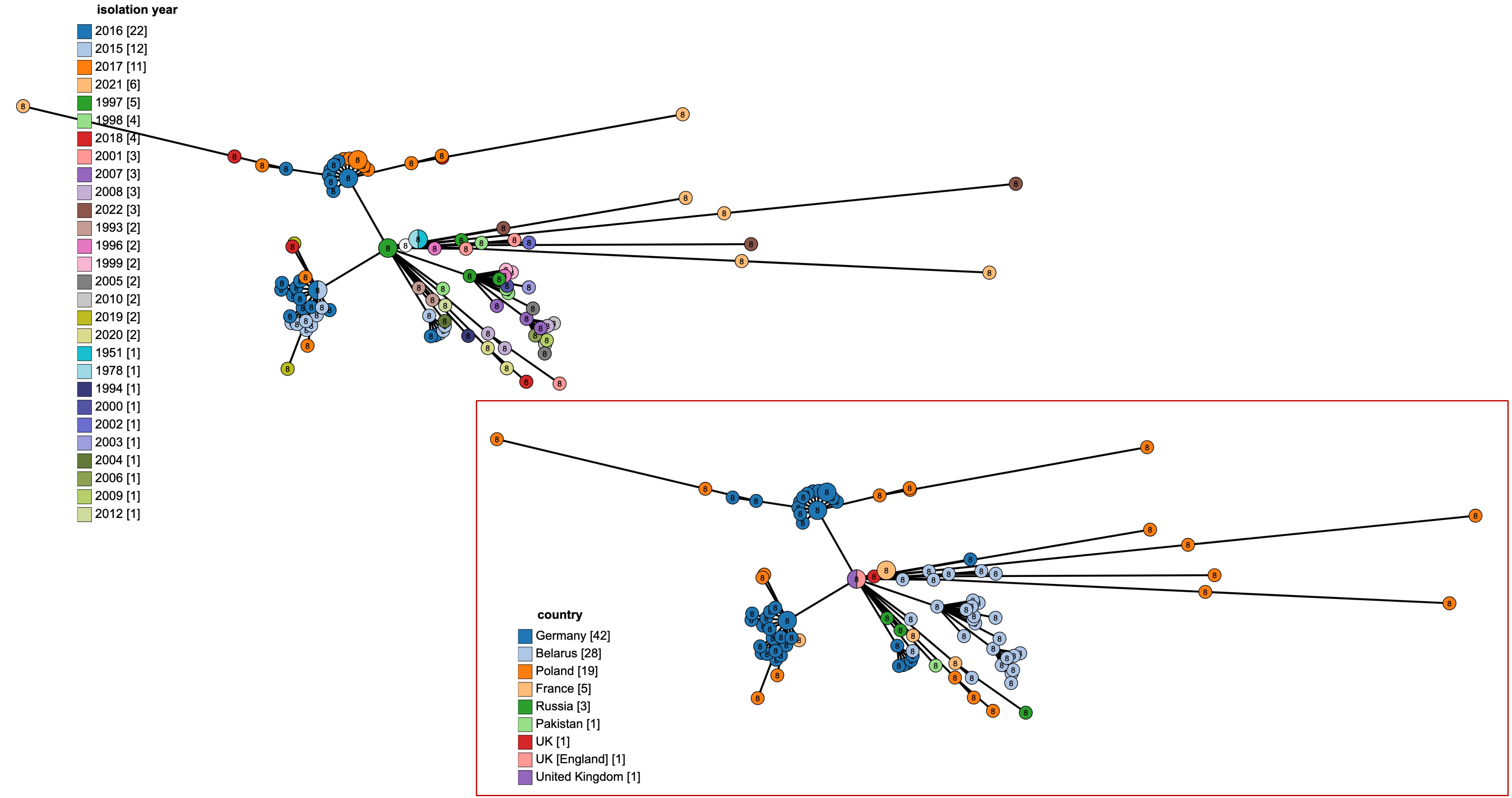
Mignon du Plessis, Rito Mikhari, Linda de Gouveia, Noluthando Duma, Tamsin Lovelock, Charlene Lawrence, Prasha Mahabeer, Yesholata Mahabeer, Nevashan Govender, Susan Nzenze, Jonathan Featherston, Mishalan Moodley, Jocelyn Moyes, Sibongile Walaza, Cheryl Cohen, Anne von Gottberg

PMID: 40023798 PMCID: PMC11878320 DOI: 10.3201/eid3103.241211

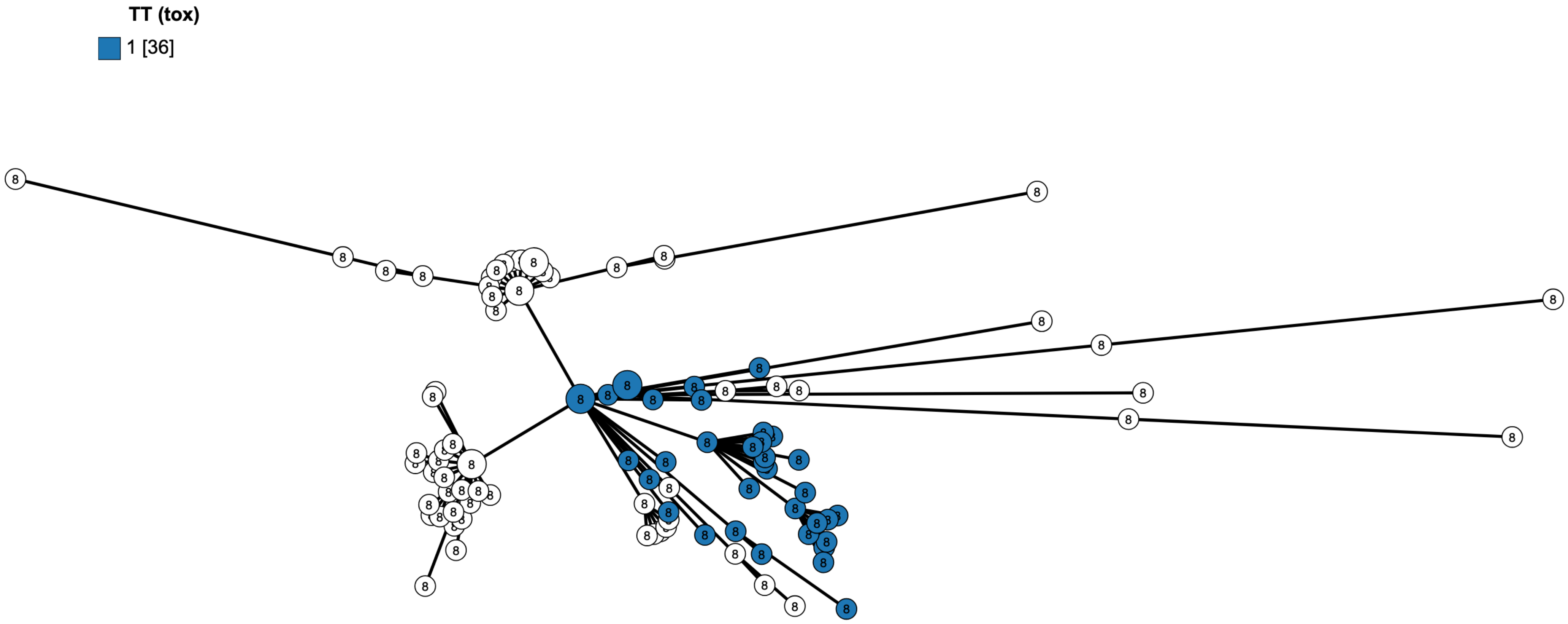
CdSC (*Corynebacterium diphtheriae* species complex) BIGSdb-Pasteur platform

- Search for public isolates of ST8
- How are they distributed by year, by country?
- Are all ST8 isolates *tox*-positive, or what proportion are?
- How is the cgMLST diversity of ST8 isolates distributed by tox?
- Are cgMLST Minimum Spanning trees reliable for phylogenetic relationships?

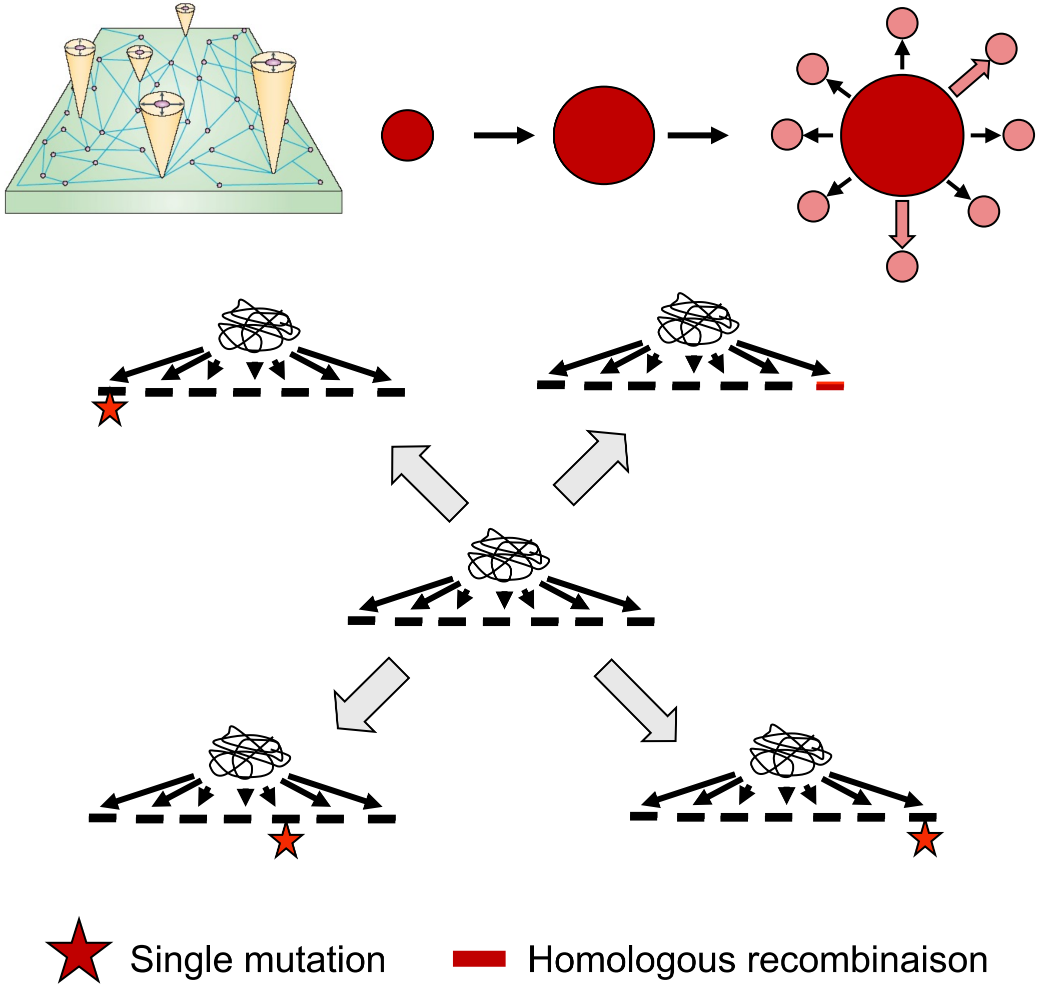
ST8 cgMLST MStree

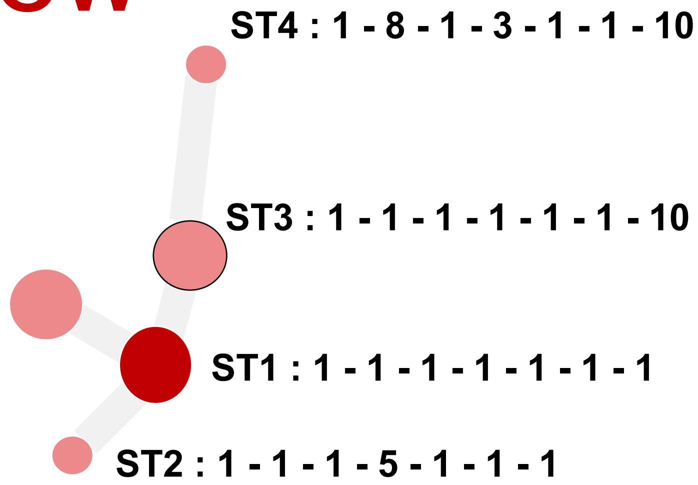


ST8 cgMLST MStree



Clonal Expansion/diversification
An MLST view



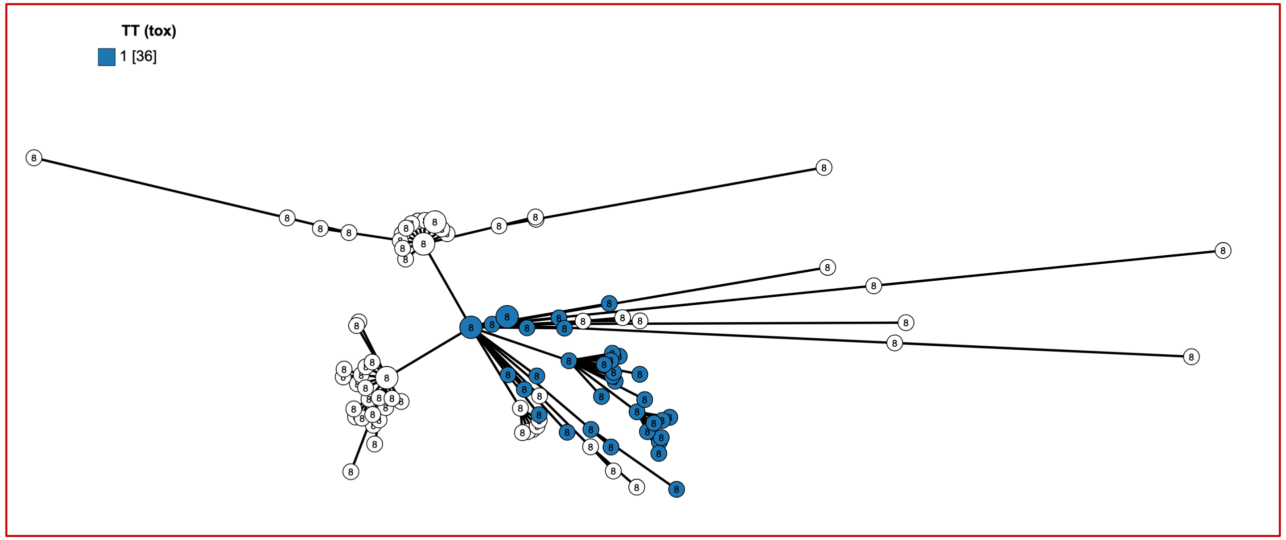


ST4 : 1 - 8 - 1 - 3 - 1 - 1 - 10
ST3 : 1 - 1 - 1 - 1 - 1 - 1 - 10
ST1 : 1 - 1 - 1 - 1 - 1 - 1 - 1
ST2 : 1 - 1 - 1 - 5 - 1 - 1 - 1

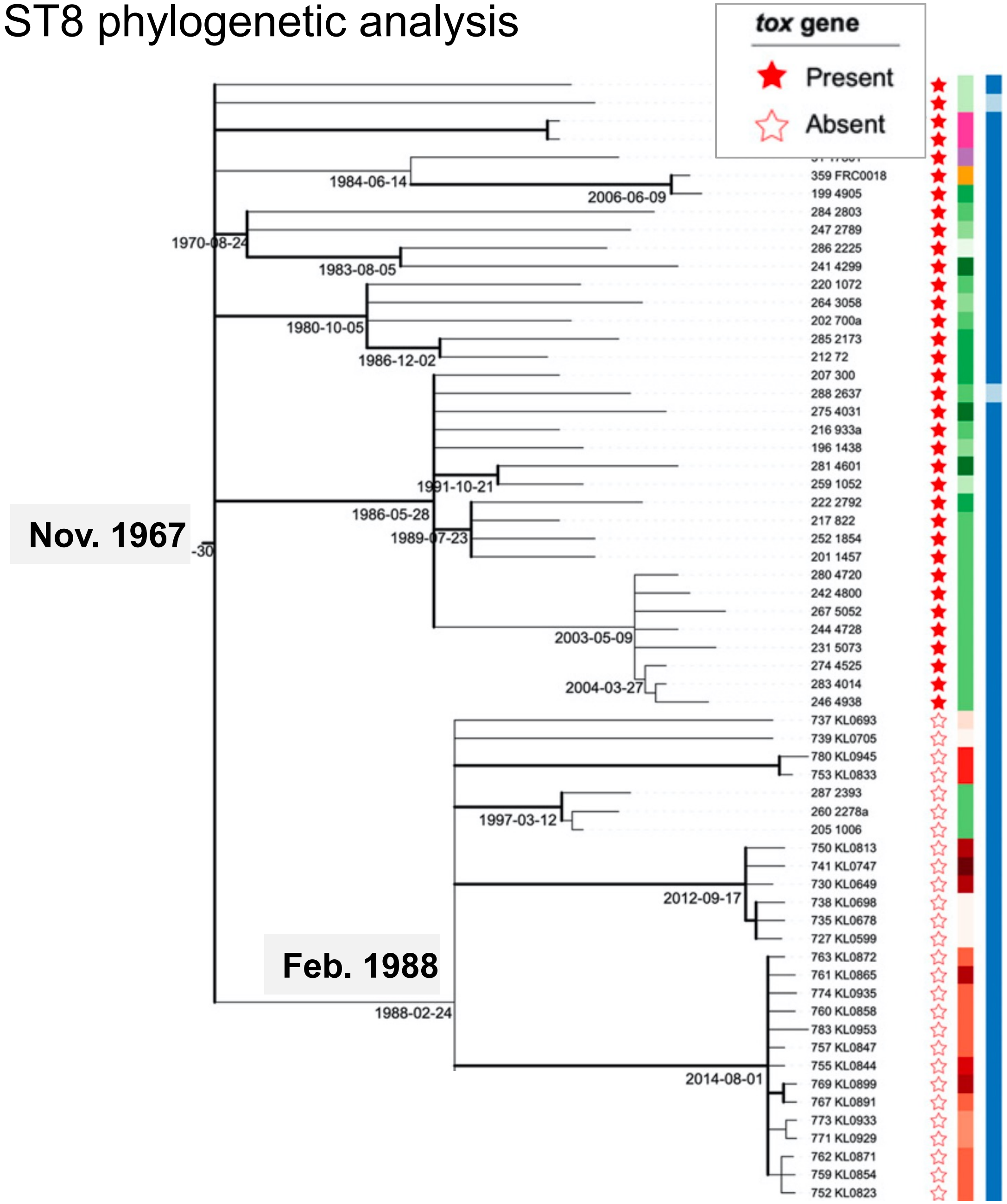
Single-linkage clustering

Classification:
Clonal complexes,
Clonal groups,…

ST8 cgMLST MStree



ST8 phylogenetic analysis



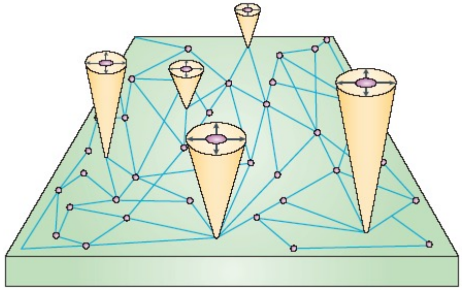
- Minimum spanning tree: hypothesizes that ancestral nodes are sampled
- Best domain of application: clonal expansion and diversification (short-term relationships with dense sampling)

Limitations of MStrees

- Typically from cgMLST data: based on raw allelic mismatch distances
- Not a phylogenetic method: Restricted evolutionary model (minimal evolution, all allelic mismatches treated as equal)
- Hypothesis: ancestral nodes are sampled:

  Only adequate at microevolutionary scales with dense sampling
  Useful from exploration of short-distance evolutionary links
  (clonal expansion-diversification model)

➡ MStrees: mostly data exploratory & visualisation tools

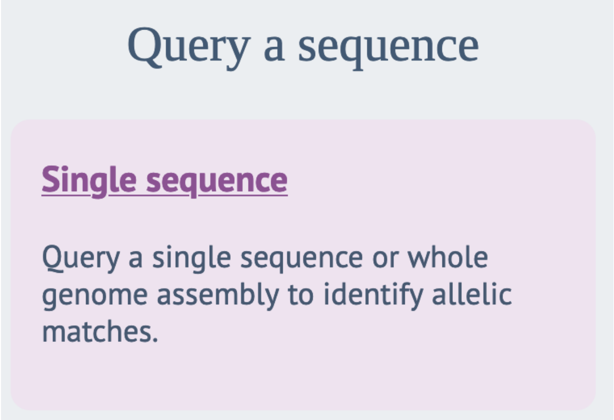


Feil, 2004

CdSC (*Corynebacterium diphtheriae* species complex) BIGSdb-Pasteur platform

- What is the ST of genome assembly Assembly1.fas?
- To what strain does this genome probably correspond to?
- Are there other related strains currently in circulation?
- What would you conclude regarding diphtheria vaccination?

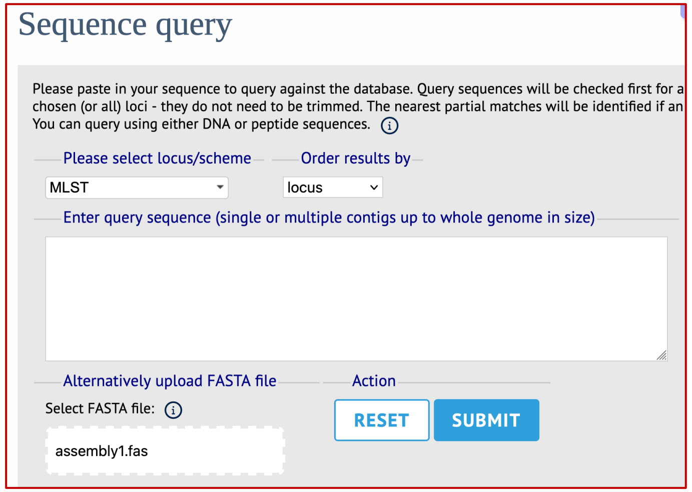
Alleles and profiles database



Query a sequence

Single sequence

Query a single sequence or whole genome assembly to identify allelic matches.



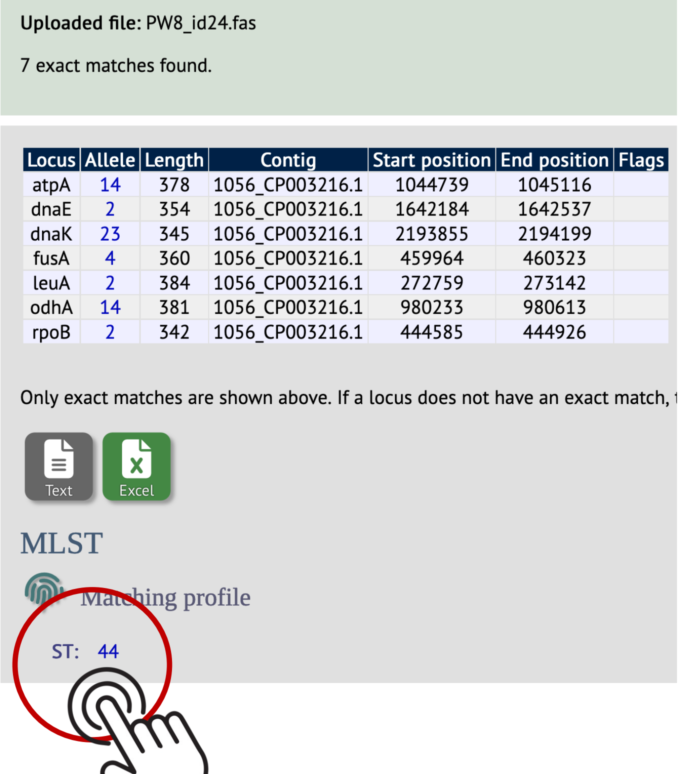
Sequence query

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

Please select locus/scheme: MLST

Order results by: locus

Select FASTA file: assembly1.fas



Uploaded file: PW8_id24.fas

7 exact matches found.

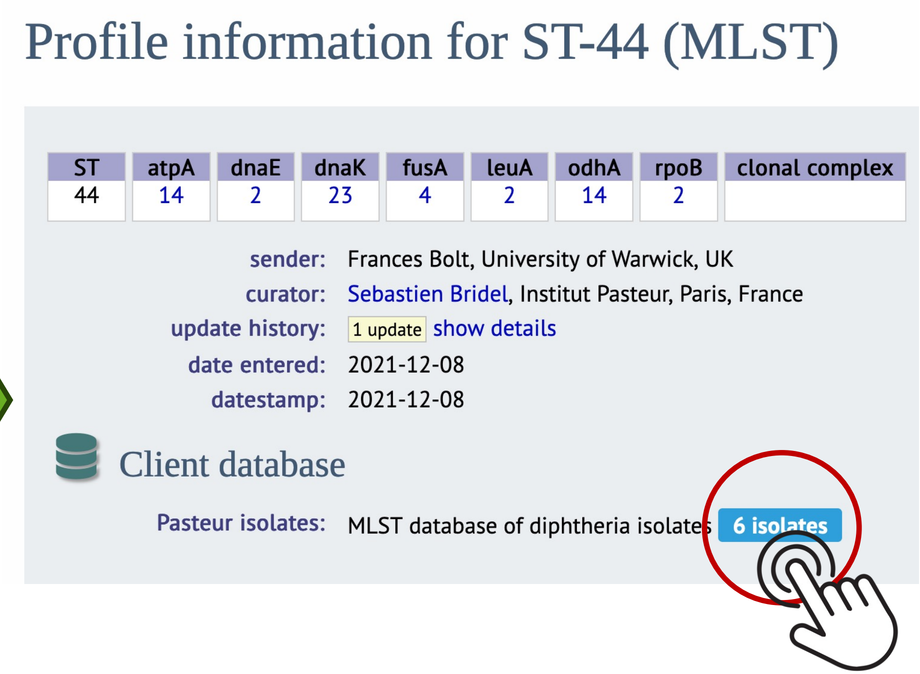
Locus	Allele	Length	Contig	Start position	End position	Flags
atpA	14	378	1056_CP003216.1	1044739	1045116	
dnaE	2	354	1056_CP003216.1	1642184	1642537	
dnaK	23	345	1056_CP003216.1	2193855	2194199	
fusA	4	360	1056_CP003216.1	459964	460323	
leuA	2	384	1056_CP003216.1	272759	273142	
odhA	14	381	1056_CP003216.1	980233	980613	
rpoB	2	342	1056_CP003216.1	444585	444926	

Only exact matches are shown above. If a locus does not have an exact match,

MLST

Matching profile

ST: 44



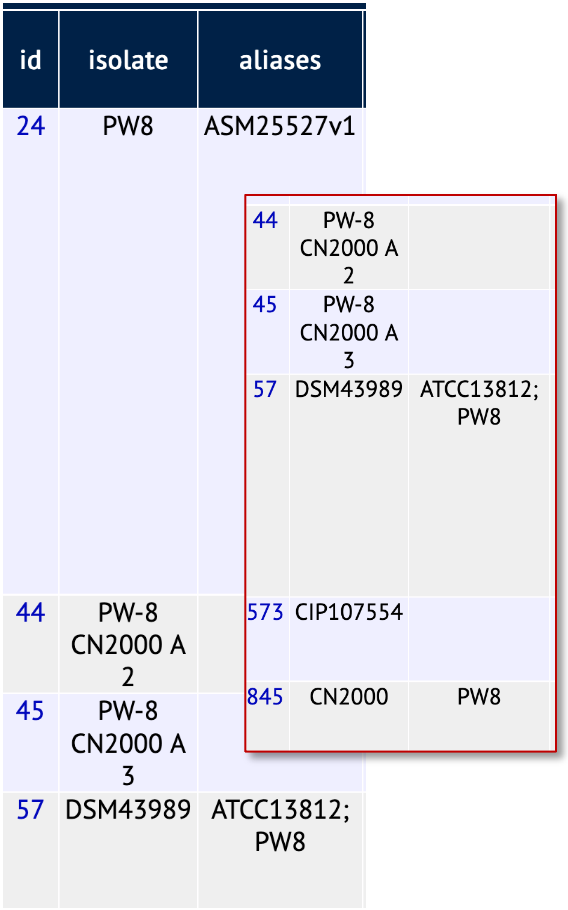
Profile information for ST-44 (MLST)

ST	atpA	dnaE	dnaK	fusA	leuA	odhA	rpoB	clonal complex
44	14	2	23	4	2	14	2	

sender: Frances Bolt, University of Warwick, UK
curator: Sebastien Bridel, Institut Pasteur, Paris, France
update history: 1 update show details
date entered: 2021-12-08
datestamp: 2021-12-08

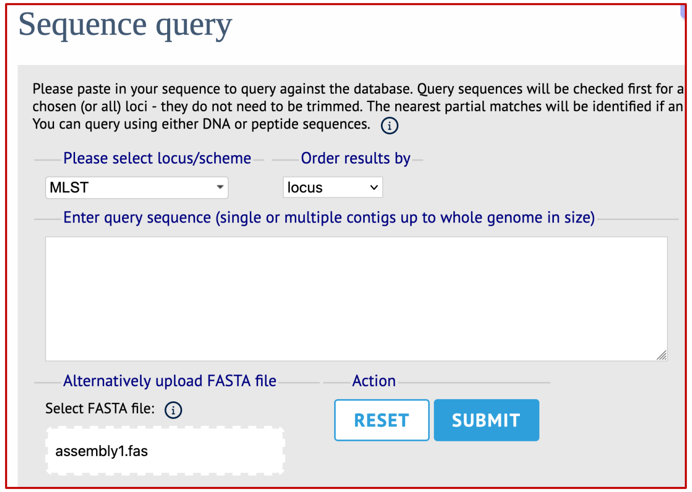
Client database

Pasteur isolates: MLST database of diphtheria isolates 6 isolates



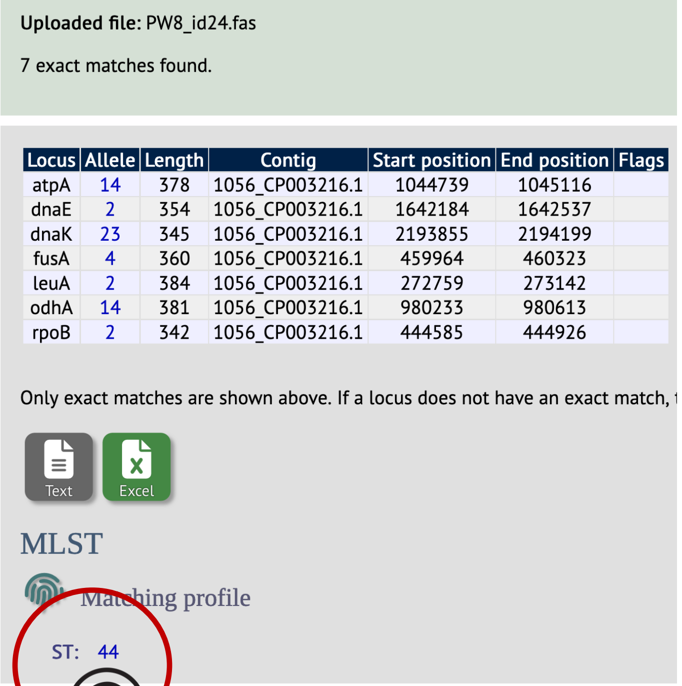
id	isolate	aliases
24	PW8	ASM25527v1
44	PW-8 CN2000 A 2	
45	PW-8 CN2000 A 3	
57	DSM43989	ATCC13812; PW8

44	PW-8 CN2000 A 2	
45	PW-8 CN2000 A 3	
57	DSM43989	ATCC13812; PW8
573	CIP107554	
845	CN2000	PW8



Sequence query

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



Uploaded file: PW8_id24.fas

7 exact matches found.

Locus	Allele	Length	Contig	Start position	End position	Flags
atpA	14	378	1056_CP003216.1	1044739	1045116	
dnaE	2	354	1056_CP003216.1	1642184	1642537	
dnaK	23	345	1056_CP003216.1	2193855	2194199	
fusA	4	360	1056_CP003216.1	459964	460323	
leuA	2	384	1056_CP003216.1	272759	273142	
odhA	14	381	1056_CP003216.1	980233	980613	
rpoB	2	342	1056_CP003216.1	444585	444926	

Only exact matches are shown above. If a locus does not have an exact match,

MLST

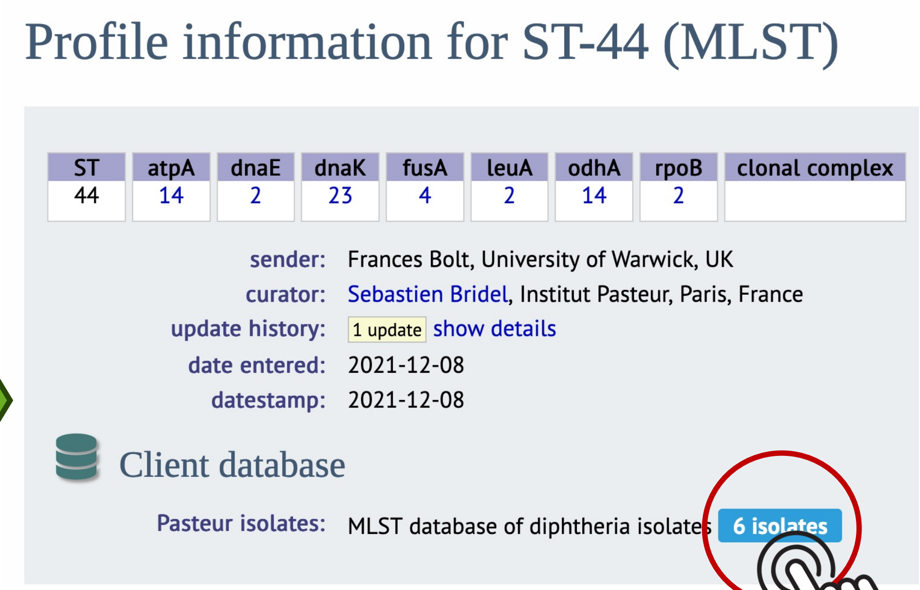
Matching profile

ST: 44

This sequence type (ST44) has not been sampled in modern times (so far)

PW8: Park-Williams 8

Vaccine strain



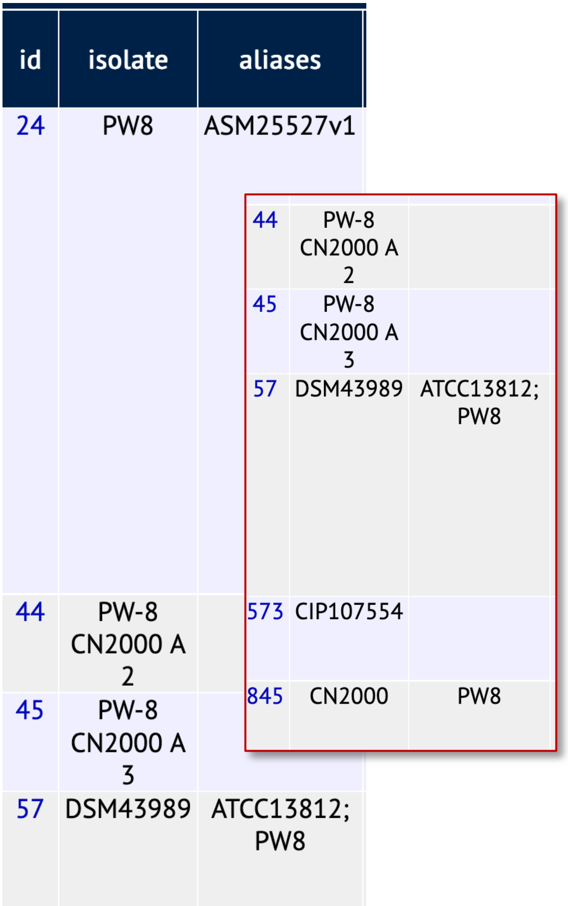
Profile information for ST-44 (MLST)

ST	atpA	dnaE	dnaK	fusA	leuA	odhA	rpoB	clonal complex
44	14	2	23	4	2	14	2	

sender: Frances Bolt, University of Warwick, UK
curator: Sebastien Bridel, Institut Pasteur, Paris, France
update history: 1 update show details
date entered: 2021-12-08
datestamp: 2021-12-08

Client database

Pasteur isolates: MLST database of diphtheria isolates 6 isolates



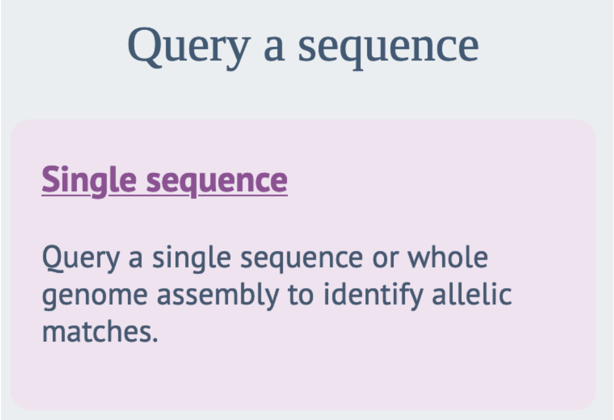
id	isolate	aliases
24	PW8	ASM25527v1
44	PW-8 CN2000 A 2	
45	PW-8 CN2000 A 3	
57	DSM43989	ATCC13812; PW8

44	PW-8 CN2000 A 2	
45	PW-8 CN2000 A 3	
57	DSM43989	ATCC13812; PW8
573	CIP107554	
845	CN2000	PW8

CdSC (*Corynebacterium diphtheriae* species complex) BIGSdb-Pasteur platform

- What is the ST of genome assembly assembly2.fas?
- To what strain does this genome probably correspond to?
- Are there other related strains currently in circulation?

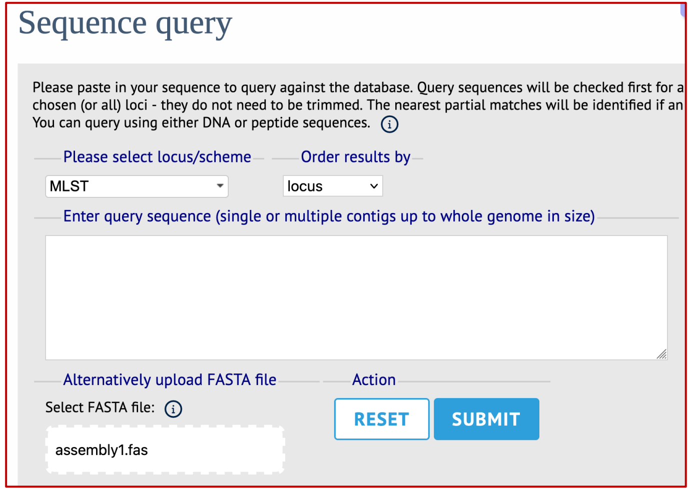
Alleles and profiles database



Query a sequence

Single sequence

Query a single sequence or whole genome assembly to identify allelic matches.



Sequence query

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

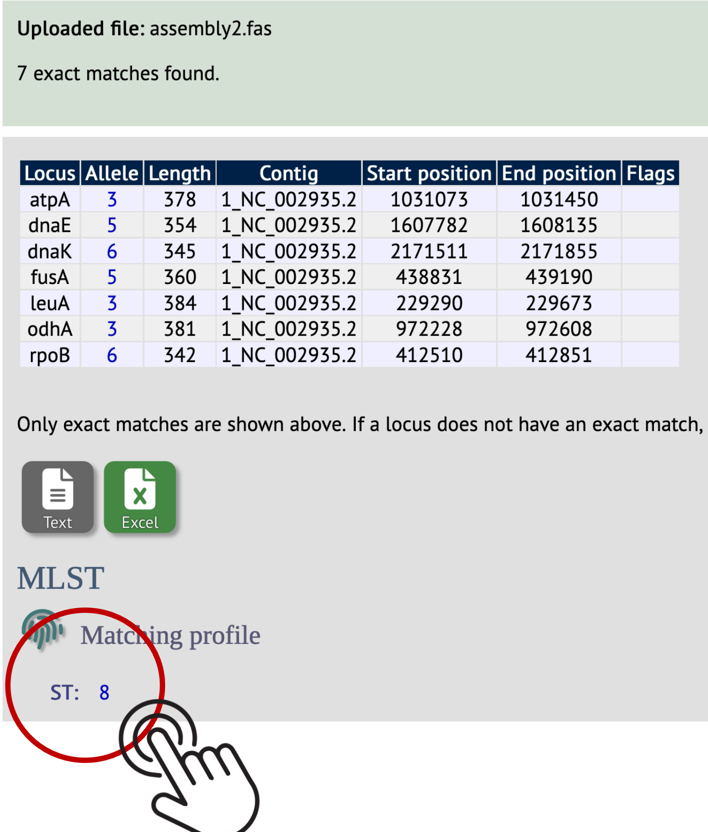
Please select locus/scheme: MLST

Order results by: locus

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

Alternatively upload FASTA file

Select FASTA file: assembly1.fas



Uploaded file: assembly2.fas

7 exact matches found.

Locus	Allele	Length	Contig	Start position	End position	Flags
atpA	3	378	1_NC_002935.2	1031073	1031450	
dnaE	5	354	1_NC_002935.2	1607782	1608135	
dnaK	6	345	1_NC_002935.2	2171511	2171855	
fusA	5	360	1_NC_002935.2	438831	439190	
leuA	3	384	1_NC_002935.2	229290	229673	
odhA	3	381	1_NC_002935.2	972228	972608	
rpoB	6	342	1_NC_002935.2	412510	412851	

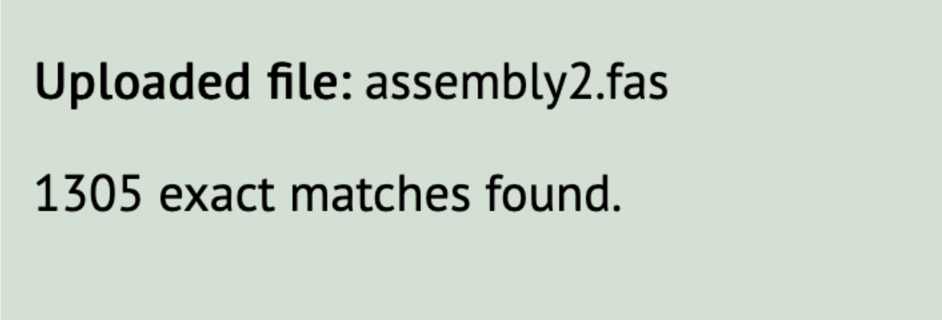
Only exact matches are shown above. If a locus does not have an exact match,

MLST

Matching profile

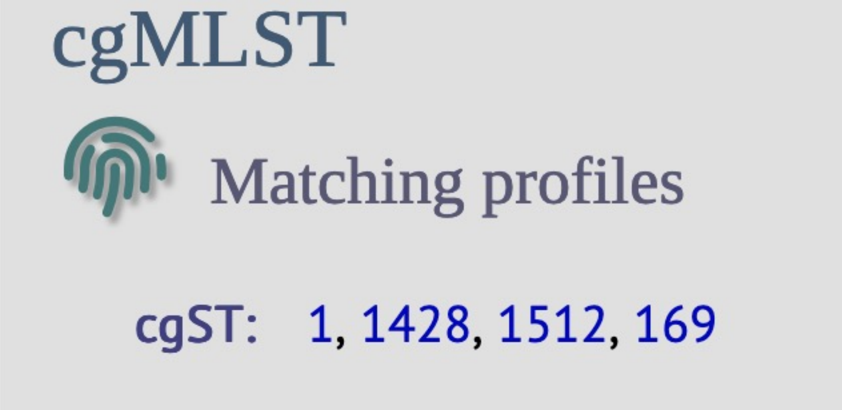
ST: 8

Scanning assembly2.fas for cgMLST



Uploaded file: assembly2.fas

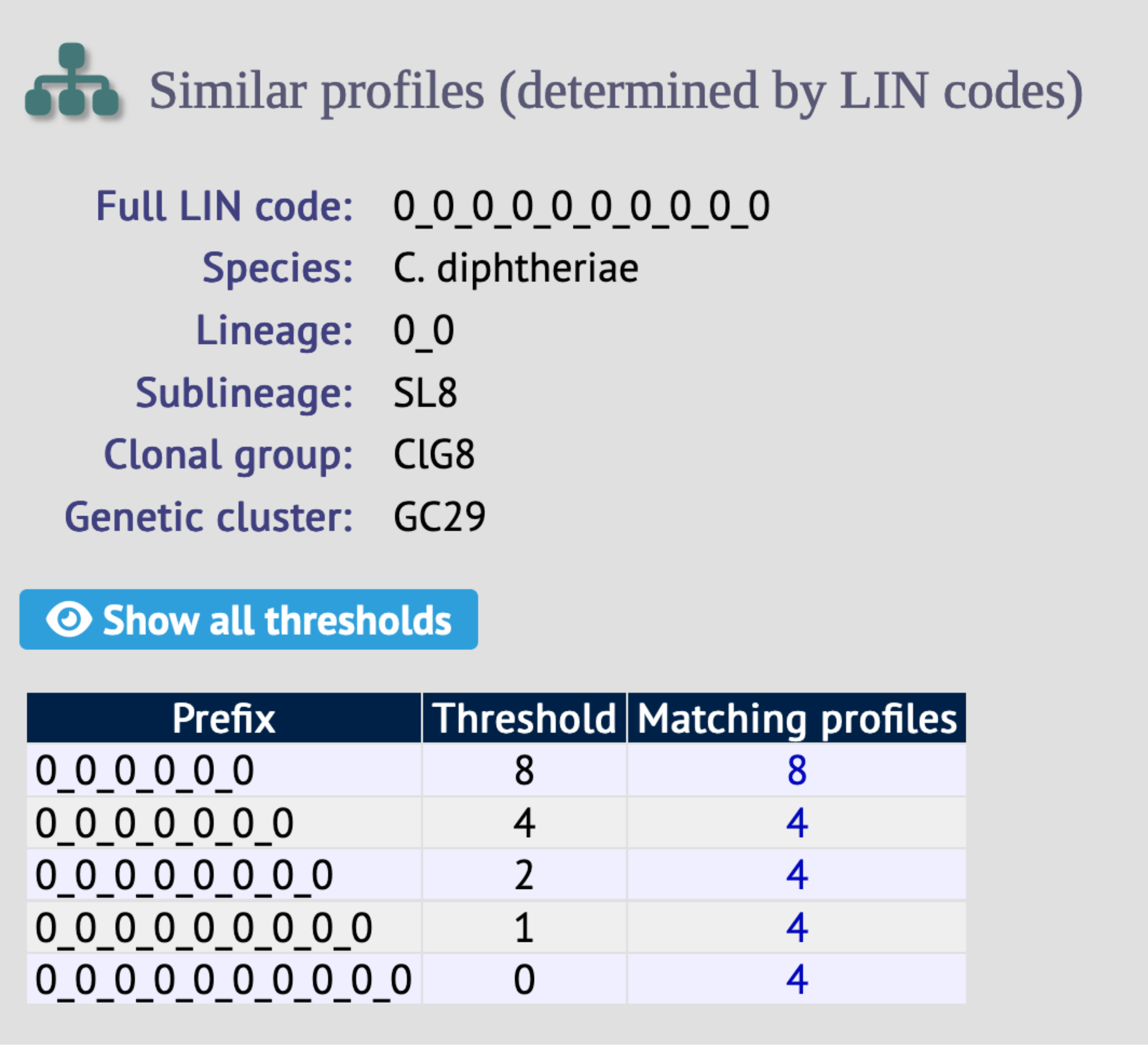
1305 exact matches found.



cgMLST

Matching profiles

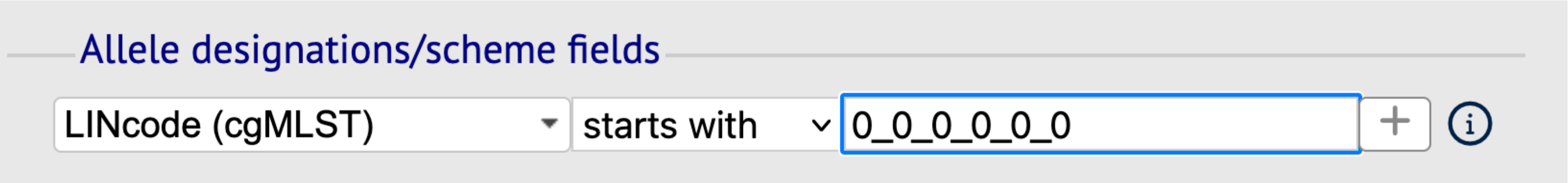
cgST: 1, 1428, 1512, 169



Similar profiles (determined by LIN codes)

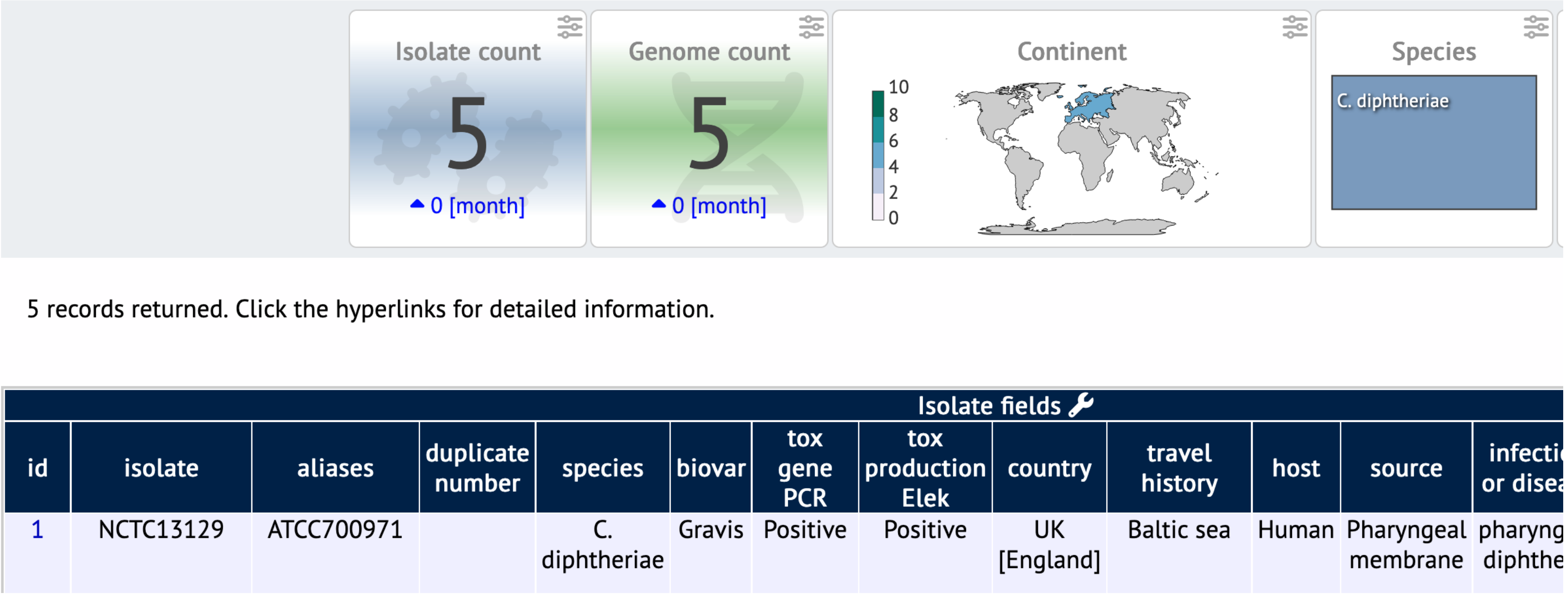
Full LIN code: 0_0_0_0_0_0_0_0_0_0
Species: C. diphtheriae
Lineage: 0_0
Sublineage: SL8
Clonal group: ClG8
Genetic cluster: GC29

Prefix	Threshold	Matching profiles
0_0_0_0_0_0	8	8
0_0_0_0_0_0_0	4	4
0_0_0_0_0_0_0_0	2	4
0_0_0_0_0_0_0_0_0	1	4
0_0_0_0_0_0_0_0_0_0	0	4



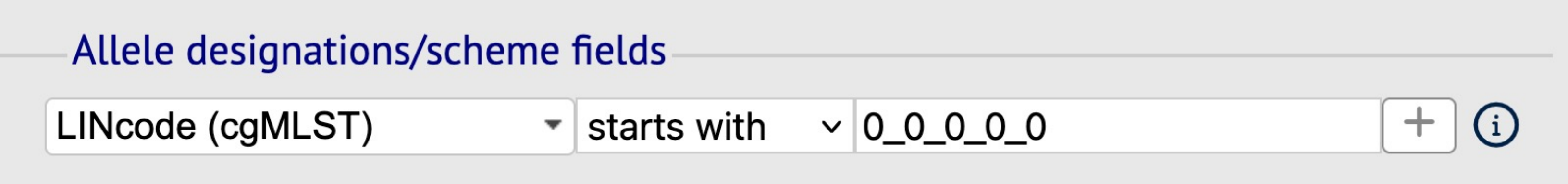
Allele designations/scheme fields

LINcode (cgMLST) | starts with | 0_0_0_0_0_0



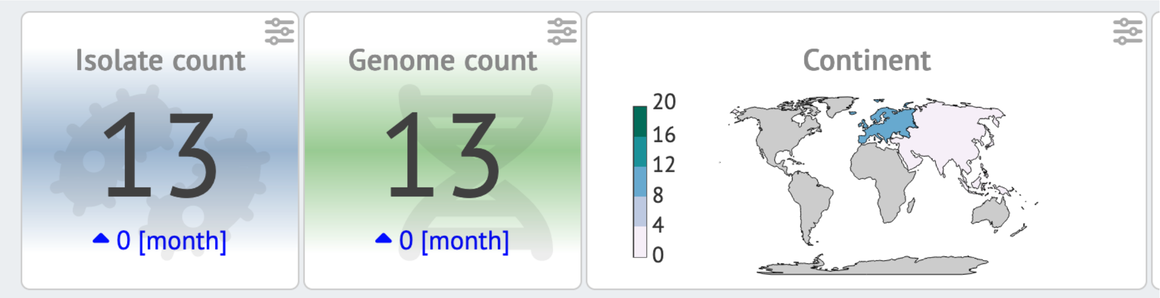
5 records returned. Click the hyperlinks for detailed information.

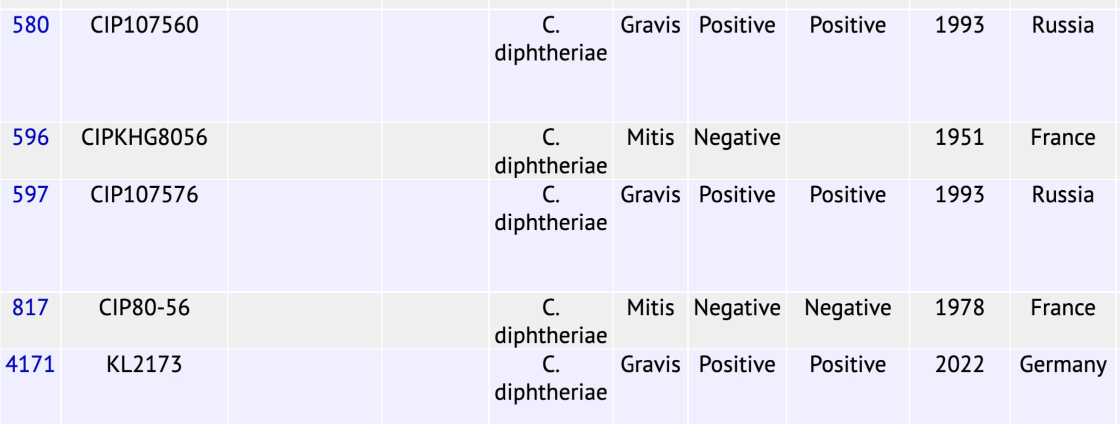
Isolate fields													
id	isolate	aliases	duplicate number	species	biovar	tox gene PCR	tox production Elek	country	travel history	host	source	infecti or disea
1	NCTC13129	ATCC700971		C. diphtheriae	Gravis	Positive	Positive	UK [England]	Baltic sea	Human	Pharyngeal membrane	pharyng diphthe



Allele designations/scheme fields

LINcode (cgMLST) | starts with | 0_0_0_0_0





580	CIP107560			C. diphtheriae	Gravis	Positive	Positive	1993	Russia
596	CIPKHG8056			C. diphtheriae	Mitis	Negative		1951	France
597	CIP107576			C. diphtheriae	Gravis	Positive	Positive	1993	Russia
817	CIP80-56			C. diphtheriae	Mitis	Negative	Negative	1978	France
4171	KL2173			C. diphtheriae	Gravis	Positive	Positive	2022	Germany

CdSC (*Corynebacterium diphtheriae* species complex) BIGSdb-Pasteur platform

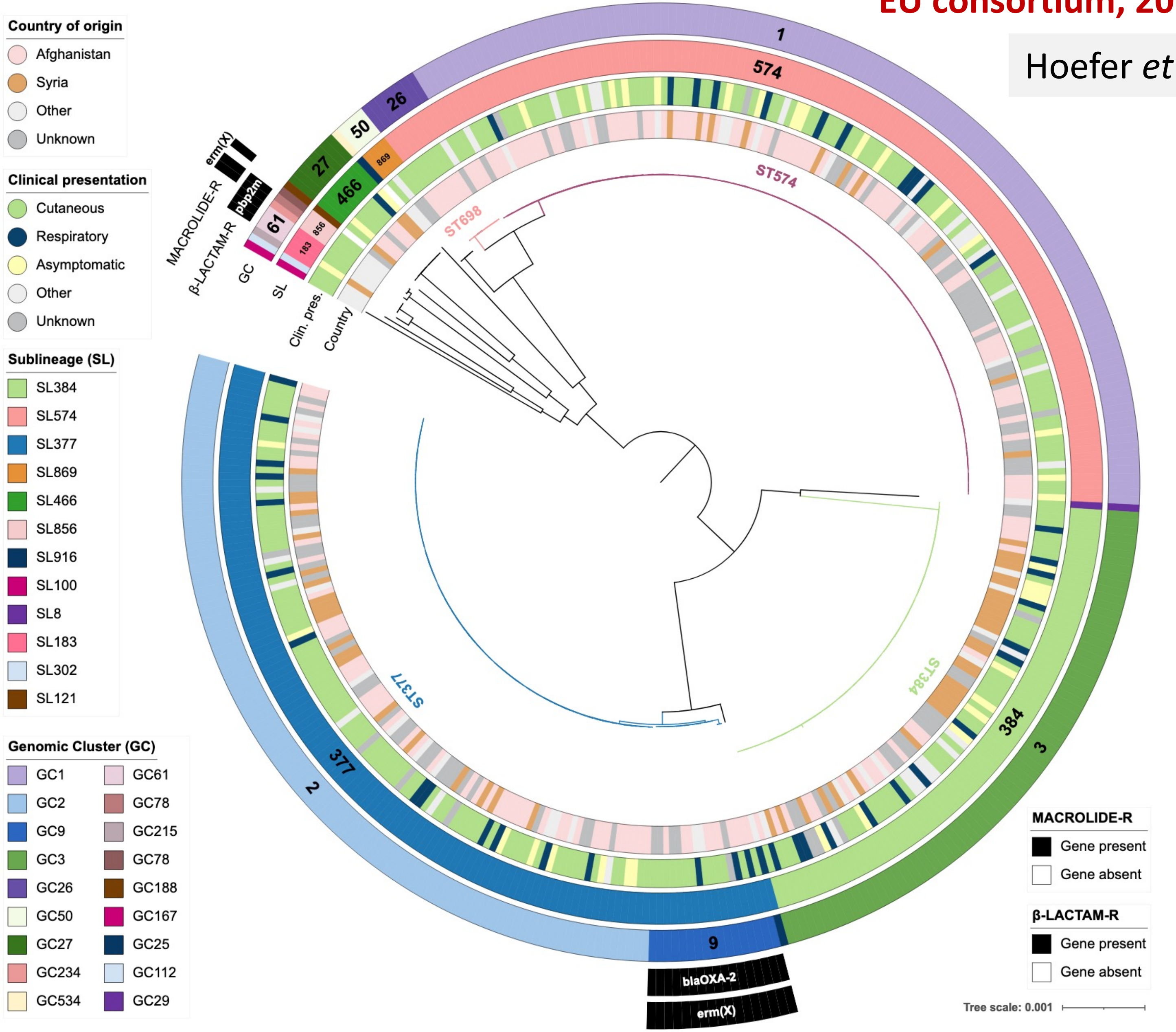
Explore public project 17: 2022_EuropeanWideReemergence_collection

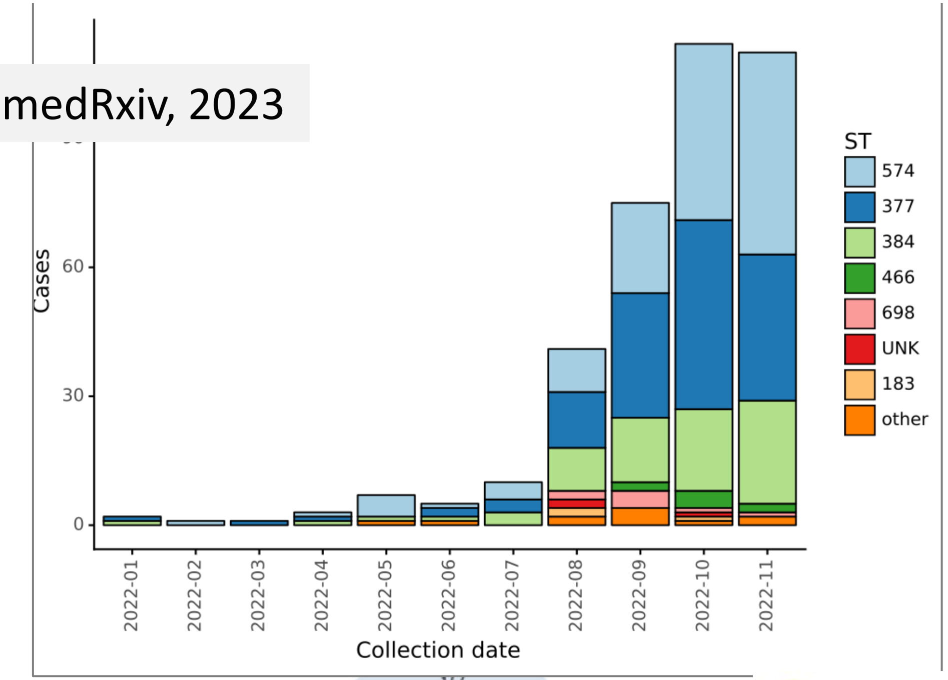
- What is its country and clinical features distribution?
- Are all isolates toxigenic, and why?
- What are its main genomic clusters and how much cgMLST diversity do they comprise?
- What would you infer regarding the emergence of these clusters?

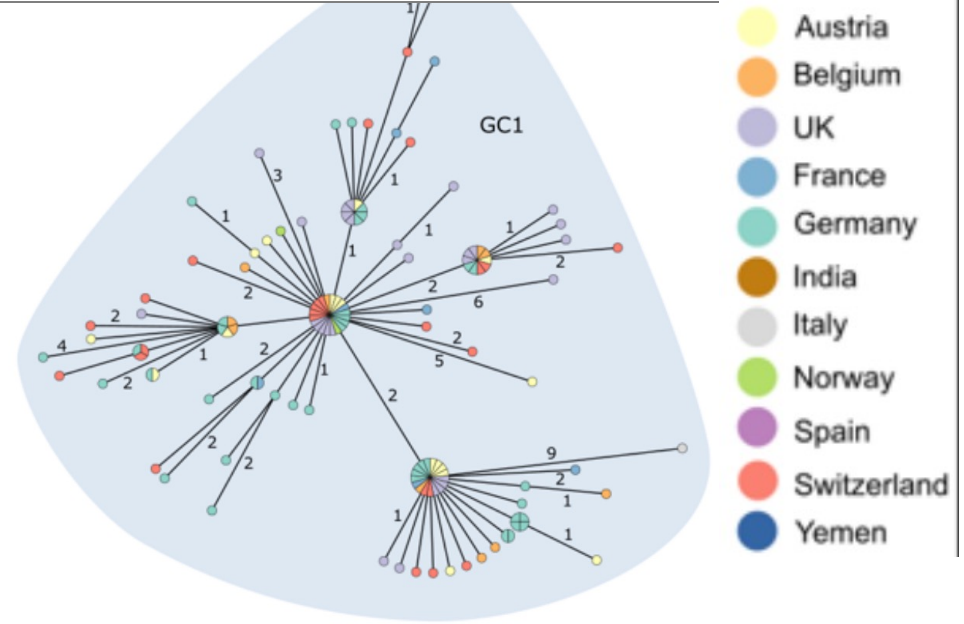
Genomics of diphtheria wave, Europe, 2022

EU consortium, 2022

Hoefer *et al.*, medRxiv, 2023







slido

Please download and install the
Slido app on all computers you use

Will you use BIGSdb for diphtheria genotyping in the near future?

ⓘ Start presenting to display the poll results on this slide.

Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of Sylvain Brisse