Virtual Training 15: Focus on the Agents of Diphtheria

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

Pr Sylvain Brisse, Institut Pasteur

Intended Learning Objectives

Learn and practice the use of BIGSdb-Pasteur platform for data integration and genomic investigation of *Corynebacterium diphtheriae*

Specific objectives of this session:

- Learn about BIGSdb contents, structure and data integration
- Learn how to explore the isolates database
- Learn how to explore the nomenclature (alleles and profiles) database
- Demonstration of a few analytical functionalities
- Presentation of practicals

slido

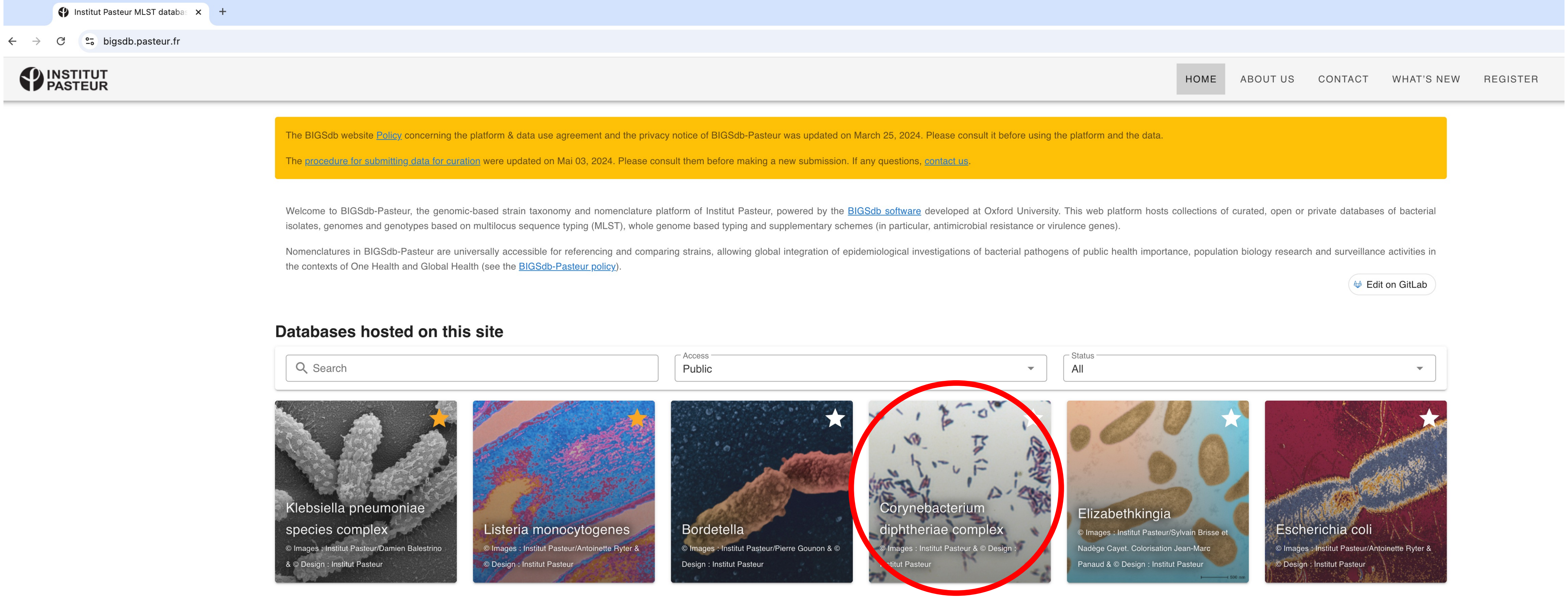
Please download and install the
Slido app on all computers you use

Have you already used the BIGSdb-Pasteur platform for diphtheria genomic analyses?

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

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

https://bigsdb.pasteur.fr/



The BIGSdb website Policy concerning the platform & data use agreement and the privacy notice of BIGSdb-Pasteur was updated on March 25, 2024. Please consult it before using the platform and the data.

The procedure for submitting data for curation were updated on Mai 03, 2024. Please consult them before making a new submission. If any questions, contact us.

Welcome to BIGSdb-Pasteur, the genomic-based strain taxonomy and nomenclature platform of Institut Pasteur, powered by the BIGSdb software developed at Oxford University. This web platform hosts collections of curated, open or private databases of bacterial isolates, genomes and genotypes based on multilocus sequence typing (MLST), whole genome based typing and supplementary schemes (in particular, antimicrobial resistance or virulence genes).

Nomenclatures in BIGSdb-Pasteur are universally accessible for referencing and comparing strains, allowing global integration of epidemiological investigations of bacterial pathogens of public health importance, population biology research and surveillance activities in the contexts of One Health and Global Health (see the BIGSdb-Pasteur policy).

Databases hosted on this site

Klebsiella pneumoniae species complex
© Images : Institut Pasteur/Damien Balestrino & © Design : Institut Pasteur

Listeria monocytogenes
© Images : Institut Pasteur/Antoinette Ryter & © Design : Institut Pasteur

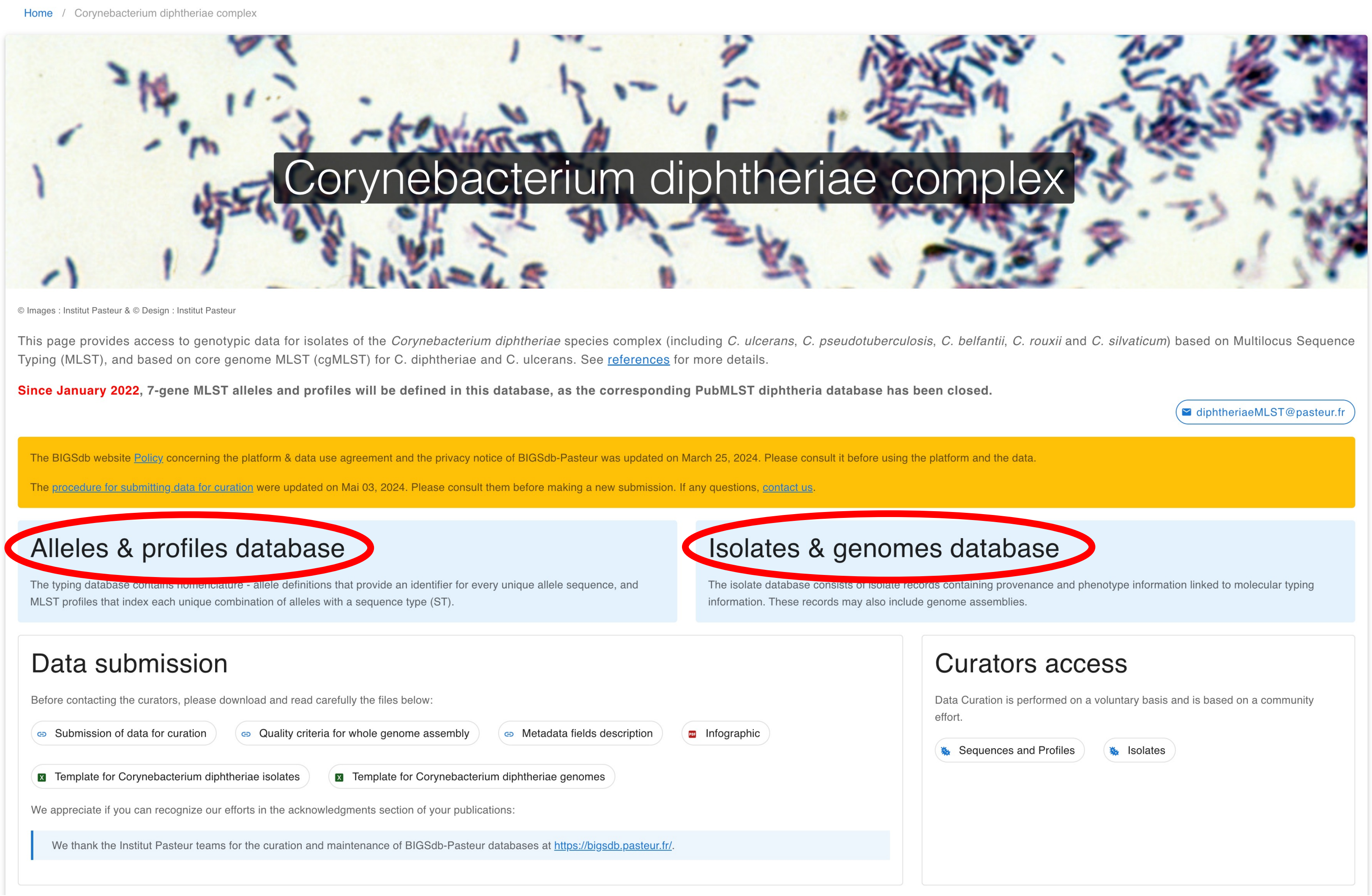
Bordetella
© Images : Institut Pasteur/Pierre Gounon & © Design : Institut Pasteur

Corynebacterium diphtheriae complex
© Images : Institut Pasteur & © Design : titut Pasteur

Elizabethkingia
© Images : Institut Pasteur/Sylvain Brisse et Nadège Cayet. Colorisation Jean-Marc Panaud & © Design : Institut Pasteur

Escherichia coli
© Images : Institut Pasteur/Antoinette Ryter & © Design : Institut Pasteur

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



Home / Corynebacterium diphtheriae complex

© Images : Institut Pasteur & © Design : Institut Pasteur

This page provides access to genotypic data for isolates of the *Corynebacterium diphtheriae* species complex (including *C. ulcerans*, *C. pseudotuberculosis*, *C. belfantii*, *C. rouxii* and *C. silvaticum*) based on Multilocus Sequence Typing (MLST), and based on core genome MLST (cgMLST) for C. diphtheriae and C. ulcerans. See references for more details.

Since January 2022, 7-gene MLST alleles and profiles will be defined in this database, as the corresponding PubMLST diphtheria database has been closed.

diphtheriaeMLST@pasteur.fr

The BIGSdb website Policy concerning the platform & data use agreement and the privacy notice of BIGSdb-Pasteur was updated on March 25, 2024. Please consult it before using the platform and the data.

The procedure for submitting data for curation were updated on Mai 03, 2024. Please consult them before making a new submission. If any questions, contact us.

Alleles & profiles database

The typing database contains nomenclature - allele definitions that provide an identifier for every unique allele sequence, and MLST profiles that index each unique combination of alleles with a sequence type (ST).

Isolates & genomes database

The isolate database consists of isolate records containing provenance and phenotype information linked to molecular typing information. These records may also include genome assemblies.

Data submission

Before contacting the curators, please download and read carefully the files below:

We appreciate if you can recognize our efforts in the acknowledgments section of your publications:

> We thank the Institut Pasteur teams for the curation and maintenance of BIGSdb-Pasteur databases at https://bigsdb.pasteur.fr/.

Curators access

Data Curation is performed on a voluntary basis and is based on a community effort.

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

Isolates database

Isolate count
3,862
▲ 0 [month]

Genome count
3,203
▲ 0 [month]

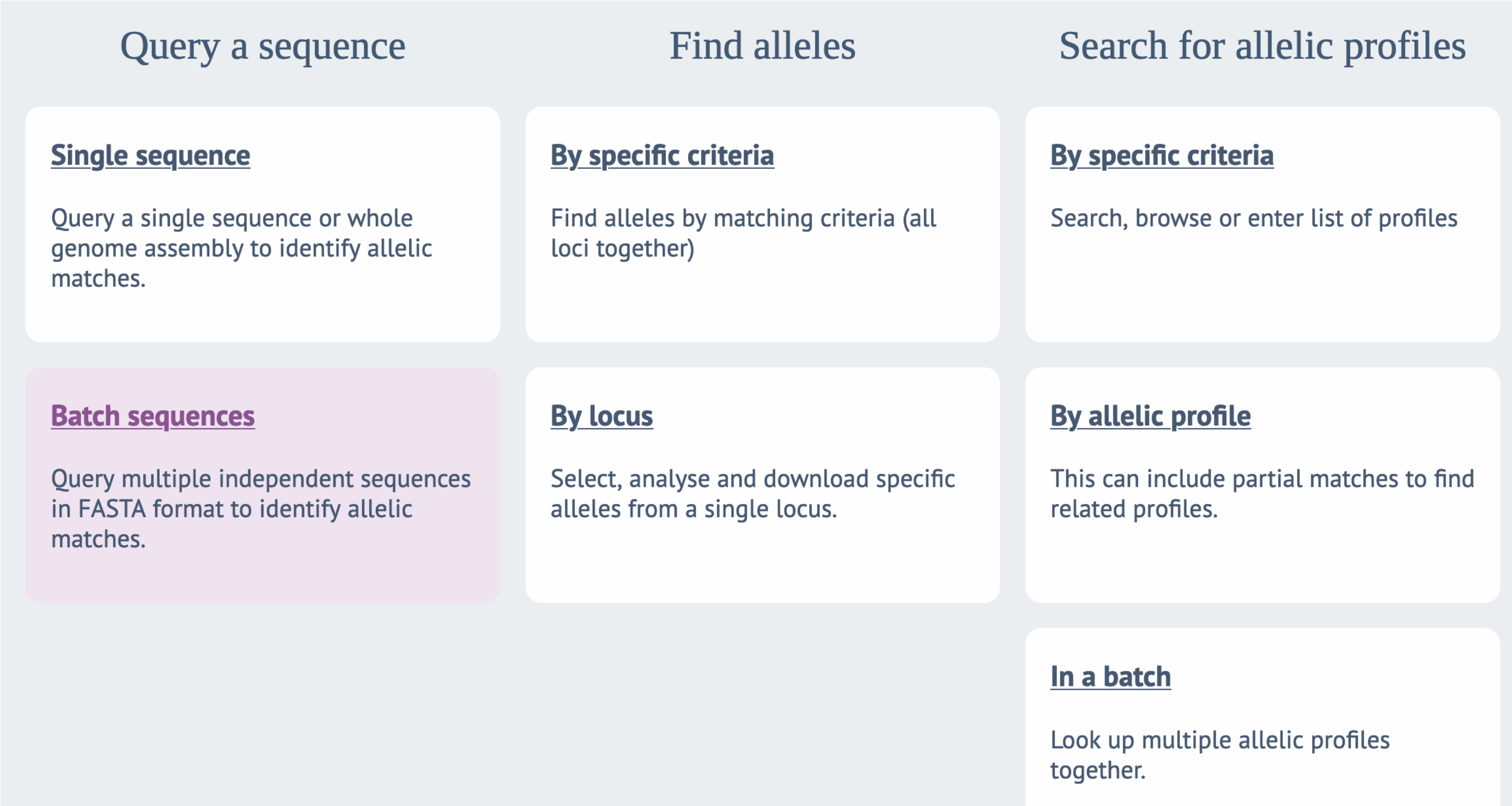
Continent
Top 5 values

Value	Frequency
Europe	2,449
Africa	328
Oceania	314
Asia	314
North America	262

No value: 922 (23.9%)

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

Alleles and profiles database



Query a sequence

Single sequence

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

Batch sequences

Query multiple independent sequences in FASTA format to identify allelic matches.

Find alleles

By specific criteria

Find alleles by matching criteria (all loci together)

By locus

Select, analyse and download specific alleles from a single locus.

Search for allelic profiles

By specific criteria

Search, browse or enter list of profiles

By allelic profile

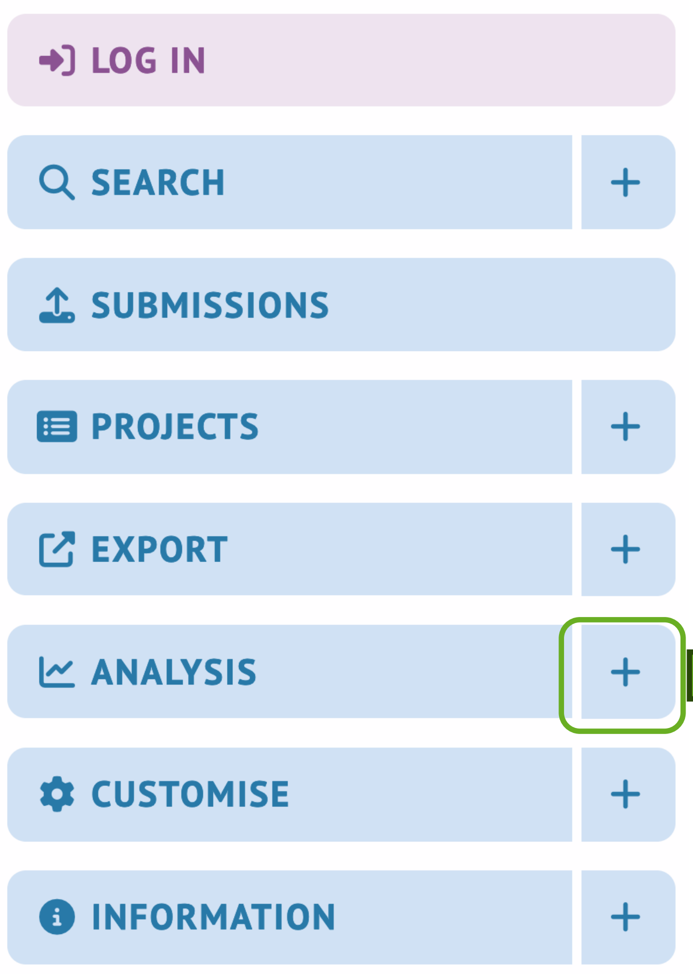
This can include partial matches to find related profiles.

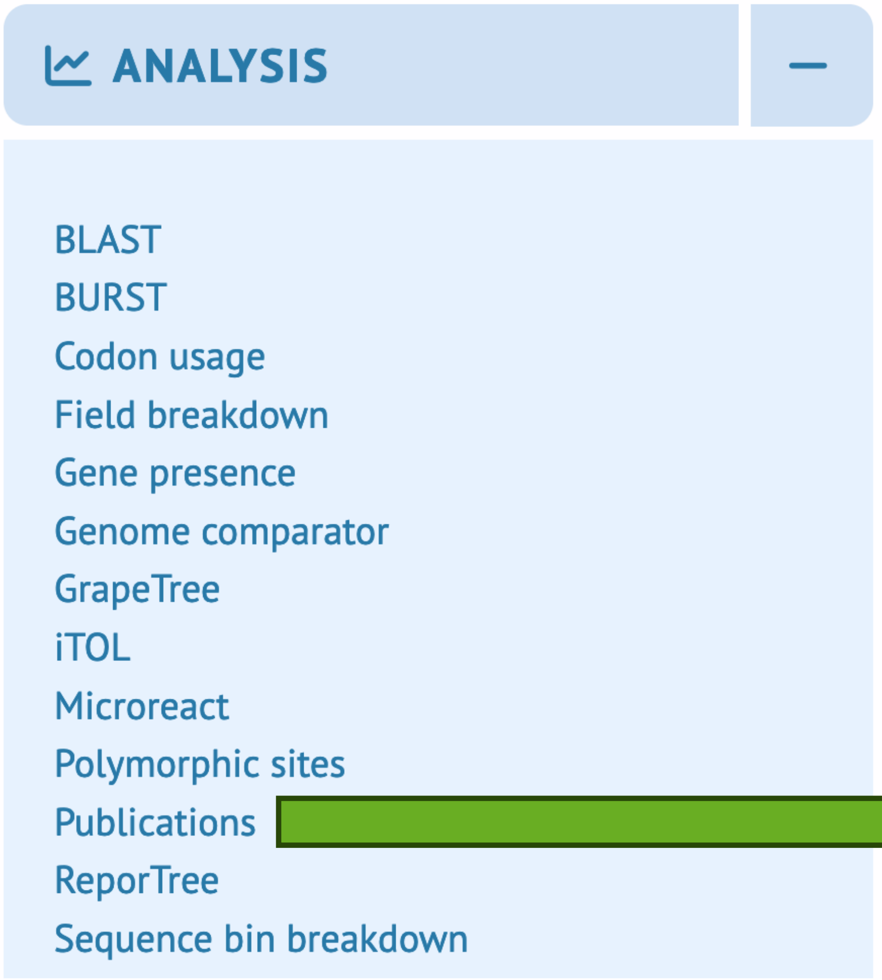
In a batch

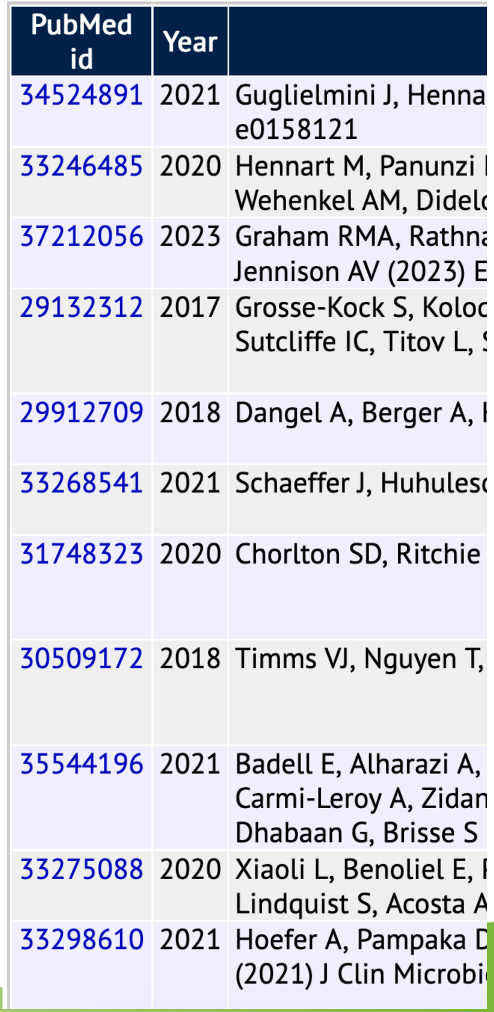
Look up multiple allelic profiles together.

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

Expandable menu items







PubMed id	Year	
34524891	2021	Guglielmini J, Henna e0158121
33246485	2020	Hennart M, Panunzi Wehenkel AM, Didel
37212056	2023	Graham RMA, Rathn Jennison AV (2023)
29132312	2017	Grosse-Kock S, Kolo Sutcliffe IC, Titov L,
29912709	2018	Dangel A, Berger A,
33268541	2021	Schaeffer J, Huhules
31748323	2020	Chorlton SD, Ritchie
30509172	2018	Timms VJ, Nguyen T,
35544196	2021	Badell E, Alharazi A, Carmi-Leroy A, Zida Dhabaan G, Brisse S
33275088	2020	Xiaoli L, Benoliel E, Lindquist S, Acosta
33298610	2021	Hoefer A, Pampaka (2021) J Clin Microb

BIGSdb consists of two main database structures

Sequence definitions

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

Allelic profiles

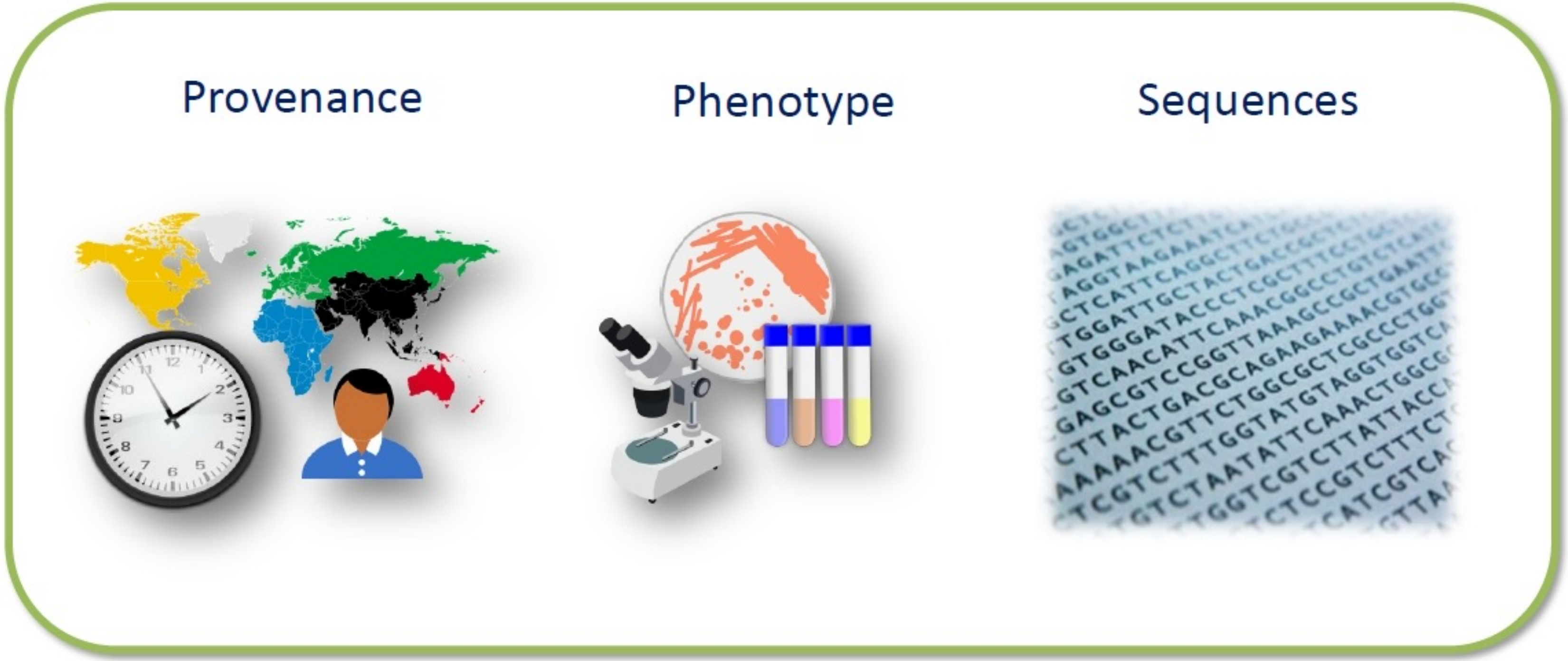
ST	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	clonal_complex
1	1	3	1	1	1	1	3	ST-1 complex
2	1	3	4	7	1	1	3	ST-1 complex
3	1	3	1	1	1	23	13	ST-1 complex
4	1	3	3	1	4	2	3	ST-4 complex
5	1	1	2	1	3	2	3	ST-5 complex
6	1	1	2	1	3	2	11	ST-5 complex
7	1	1	2	1	3	2	19	ST-5 complex
8	2	3	7	2	8	5	2	ST-8 complex
9	2	3	8	10	8	5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

(a.k.a. Alleles and profiles database)

Slide: Keith Jolley

BIGSdb consists of two main database structures



Provenance

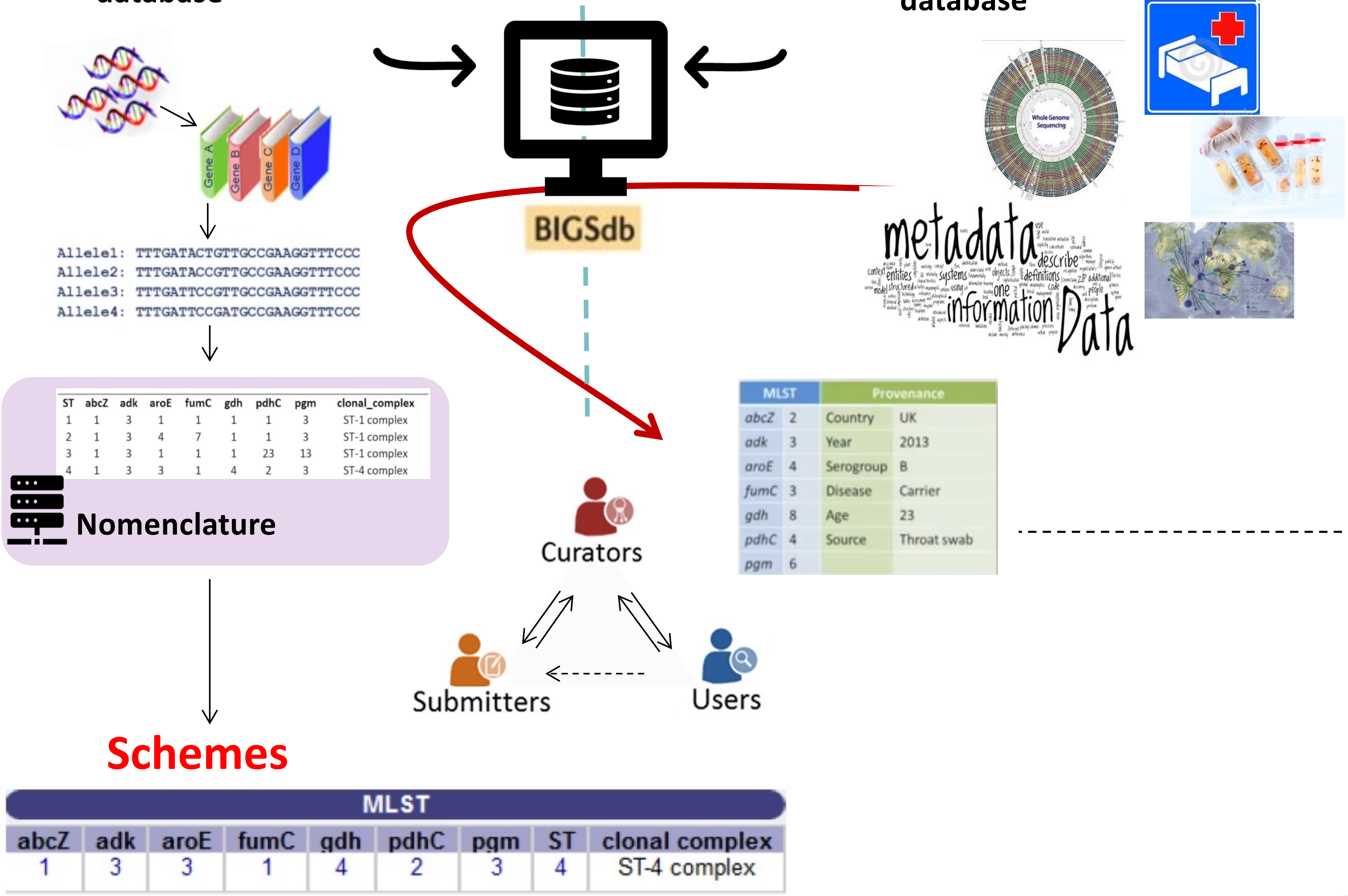
Phenotype

Sequences

Isolate data

Slide: Keith Jolley

Connectivity between isolates and nomenclature databases



Sequence & Profiles database

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

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

Nomenclature

Schemes

MLST								
abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST	clonal complex
1	3	3	1	4	2	3	4	ST-4 complex

Isolates & Genomes database

MLST		Provenance	
abcZ	2	Country	UK
adk	3	Year	2013
aroE	4	Serogroup	B
fumC	3	Disease	Carrier
gdh	8	Age	23
pdhC	4	Source	Throat swab
pgm	6		

- **Universal language**
- **Confidentiality**

International tracking
Medically relevant features
Population diversity
Epidemiosurveillance
Functional analysis
Pathoadaptation
Evolution

Live demo

Explore the isolates database section:

- Look at home page information
- Data submission: not considered today
- Main menu functions
  - **Analysis**: focus on BLAST, Microreact, publications
  - **Information**: focus on database status
  - **Search:** search database: country (e.g., USA), sublineage (e.g., SL8), LIN codes
- GrapeTree (e.g., USA MLST)

Explore allele and profiles database:

- Focus on ANALYSIS > Locus explorer > pbp2m
- Focus on ANALYSIS > Seq comparison > tox
- Sequence query and isolates lookup

slido

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

Are you intending to do the exercices before Friday?

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

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?

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?

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?

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?

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?

Acknowledgements

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