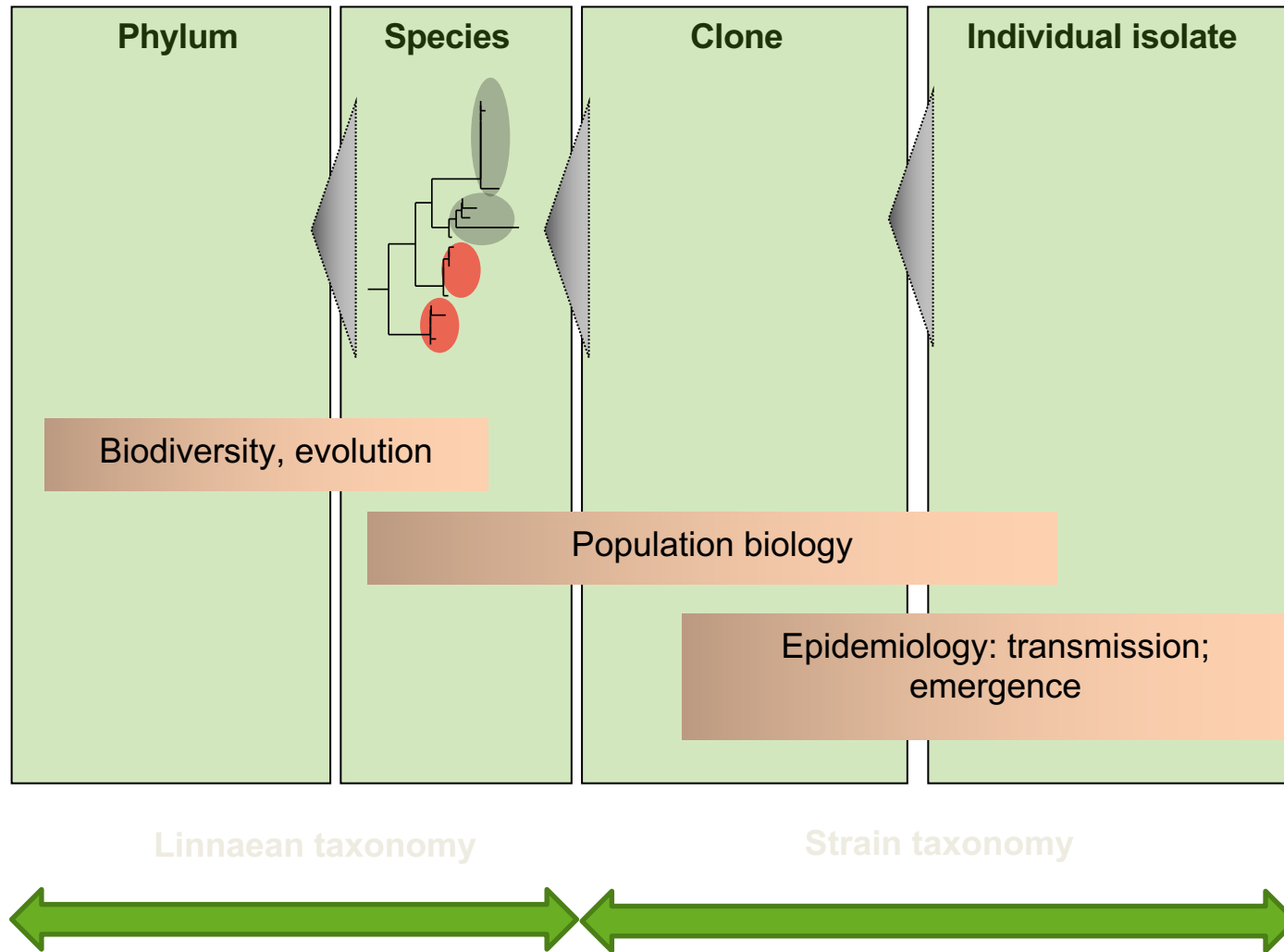




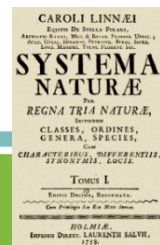
BIGSdb-Pasteur: gene-by-gene genomic epidemiology

<https://bigsdbs.pasteur.fr>

Strain taxonomies are needed, but mostly lacking



- Outbreak detection
- Epidemiology
- Genotype-phenotype links





MILST

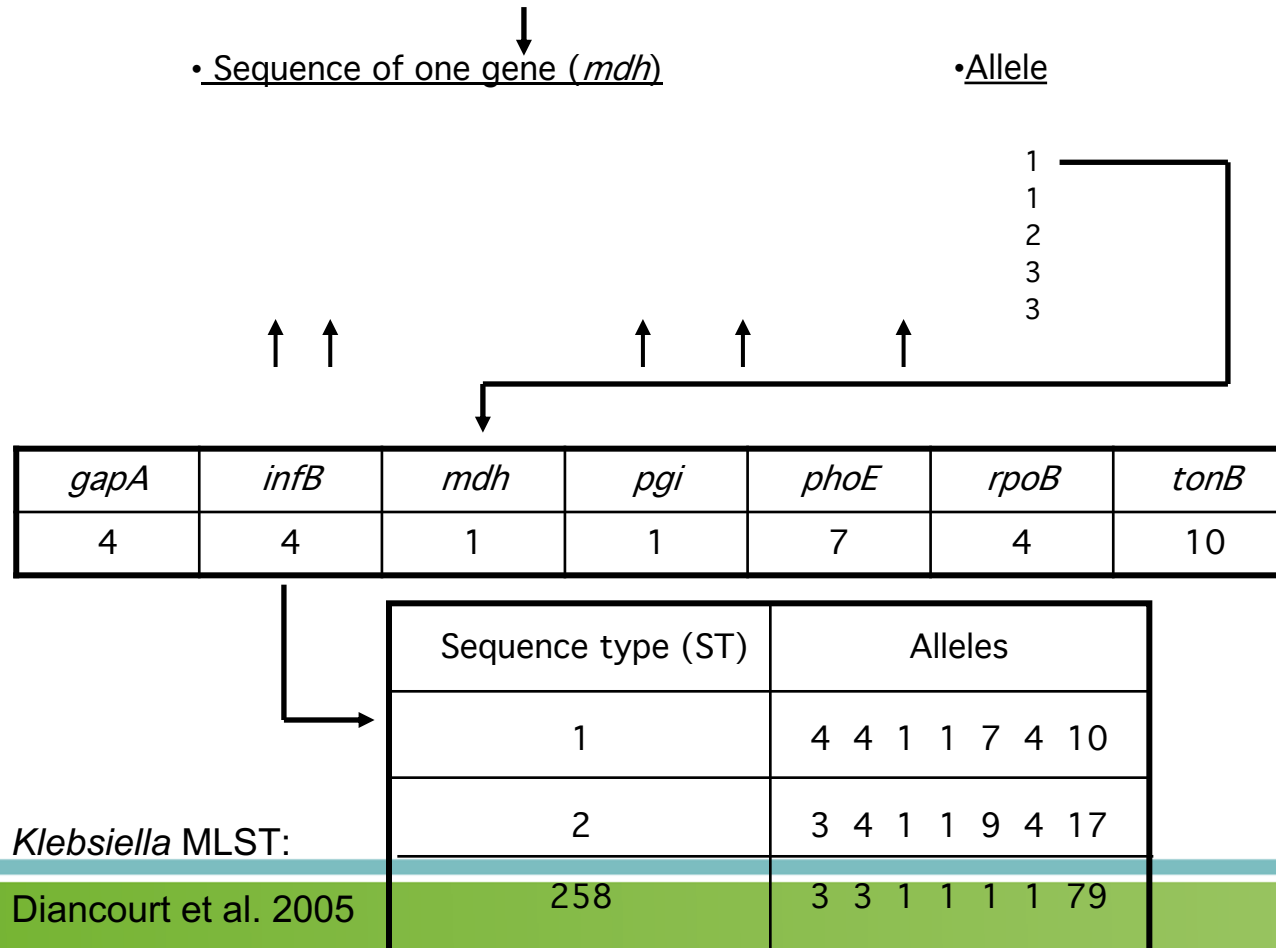
The background features a stylized illustration of a desk with a laptop, a tablet, a pen, and a coffee cup, viewed through a circular frame. The entire scene is set against a white background with a green field at the bottom and a blue line separating the field from the white area.

The Multilocus Sequence Typing standard

Bacteria
Fungi
Parasites

Maiden, Achtman, Spratt *et al.*

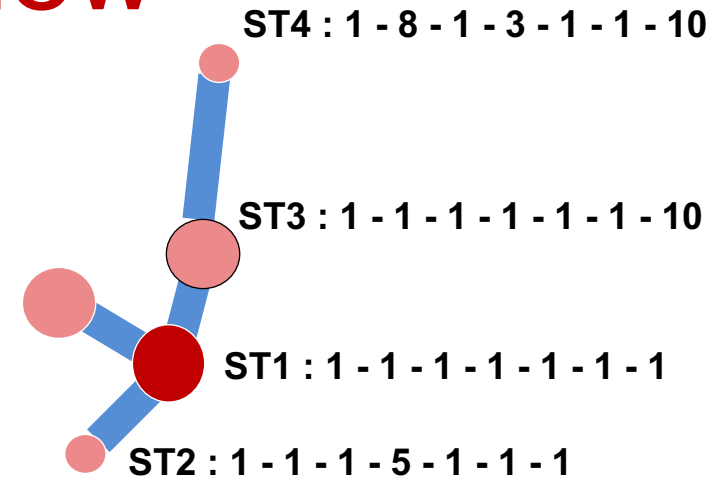
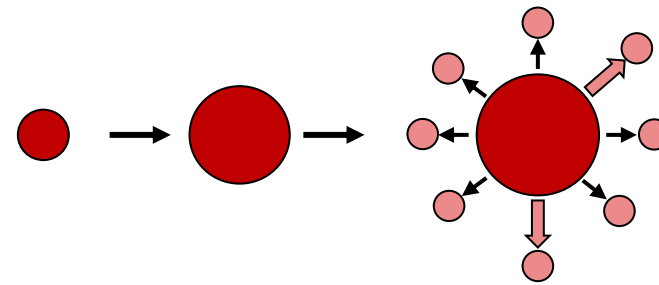
FIGURES



**Don't say *Klebsiella*,
say *Klebsiella* ST258
(or indeed SL258, if
you read Hennart et
al. 2022 Mol Biol Evol)**

Clonal Expansion/diversification

An MLST view



FIGURES

Single-linkage clustering

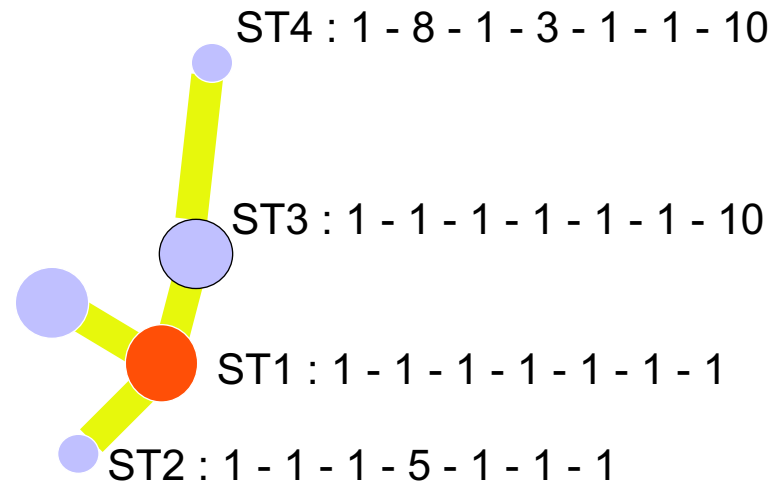
Classification:
Clonal complexes,
Clonal groups,...

★ Single mutation — Homologous recombination

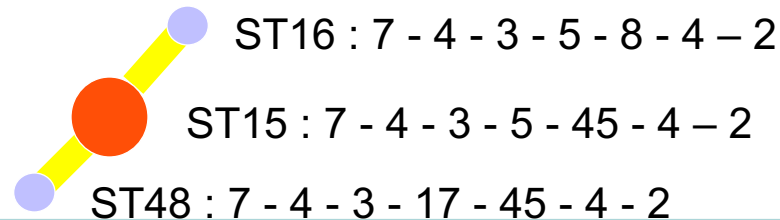
The eBURST approach

- Find closely-related genotypes and group them into clonal complexes (clonal families)
- Identify founder genotype of family

Clonal Complex 1



Clonal Complex 15



Advantages:

- Identifies founder genotype
- Shows genotype frequency
- Links STs according to distance
- Map other characteristics (e.g. serotype) onto graph as colors

eBURST analysis Minimum spanning tree



- eburst.mlst.net
- <http://goeburst.phyloviz.net/>

(e.g. software BioNumerics)



Founder genotype



- Proposes links between clonal families
- Links may be unreliable

FIGURES

Bacterial population structures



‘Minimum-spanning tree’

7-gene MLST

1605-gene cgMLST

FIGURES

99,722 *Salmonella* (Alikhan et al. 2018)

12,179 *Neisseria meningitidis* (Jolley et al. 2018)

Public MLST databases: open libraries of genotypes



pubmlst.org

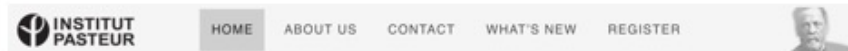
~ 100 pathogens, incl.:
Acinetobacter baumannii
Neisseria meningitidis
Campylobacter
...



enterobase.warwick.ac.uk

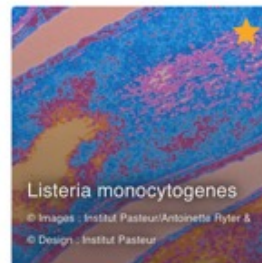
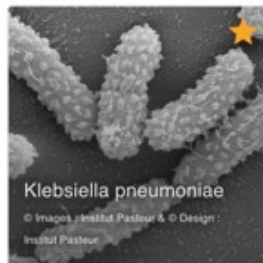
~ 7 pathogens, incl.:
Salmonella
Escherichia/Shigella
Vibrio
Yersinia
C. difficile

bigsdB.pasteur.fr



Databases hosted on this site

Search Access: Public



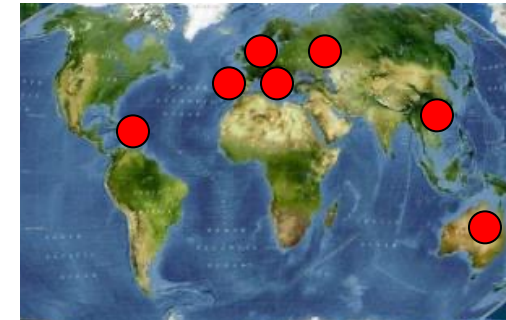
~ 10 pathogens, incl.:
Klebsiella pneumoniae
Listeria monocytogenes
Corynebacterium diphtheriae
Bordetella pertussis

<http://bigsdB.pasteur.fr/klebsiella>



[Sequences and profiles database](#)
[Isolates database](#)

- 6500 STs
- 50000 isolates
- 400 submitters



Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*

Tracking microevolution events among ST11 carbapenemase-producing hypervirulent *Klebsiella pneumoniae* outbreak strains

Common language

BIGSdb-Pasteur: gene-by-gene genomic epidemiology



PASTEUR MLST



<https://bigsdb.pasteur.fr>

Diancourt *et al.*, 2005

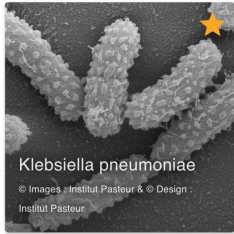
Bialek-Davenet *et al.*, 2014

Hennart *et al.* 2022 Mol Biol Evol

Hennart, Palma *et al* bioRxiv 2023

- MLST definitions (7-gene classical scheme)
- cgMLST definitions (634 loci)
- Virulence genes
- Resistance genes (blaOKP, blaLEN, blaOXY beta-lactamases)
- *wzi* (capsular typing)

7-gene MLST (*K. pneumoniae* species complex)

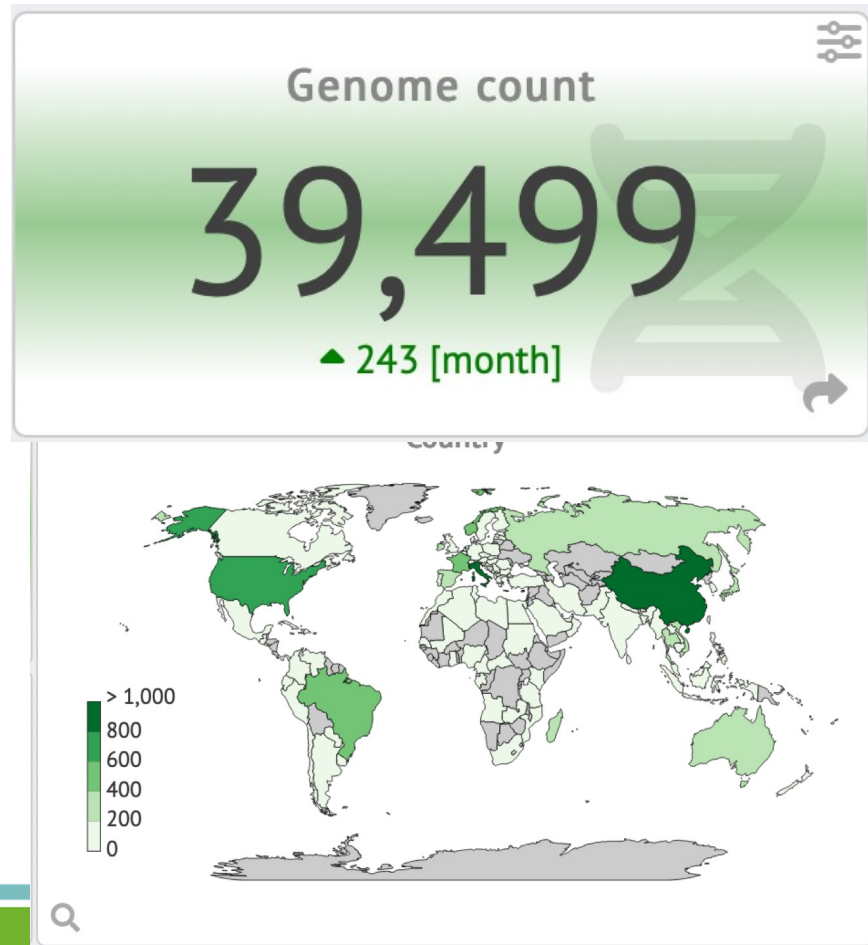


<http://www.pasteur.fr/mlst>

=

<https://bigsdb.pasteur.fr/>

Diancourt *et al.* 2005 J Clin Microbiol



- > 6500 STs
- > 40 000 public isolates
- > 5000 private isolates
- > 500 submitters

Tracking microevolution events among ST11 carbapenemase-producing hypervirulent *Klebsiella pneumoniae* outbreak strains



Widely used strain nomenclature

BIGSdb consists of two main database structures

Sequence definitions

Allelic profiles

Nomenclature

Slide: Keith Jolley

BIGSdb consists of two main database structures

Provenance

Phenotype

Sequences

Isolate data

Slide: Keith Jolley

Databases of genomes and isolates metadata applications



FIGURES

Schemes

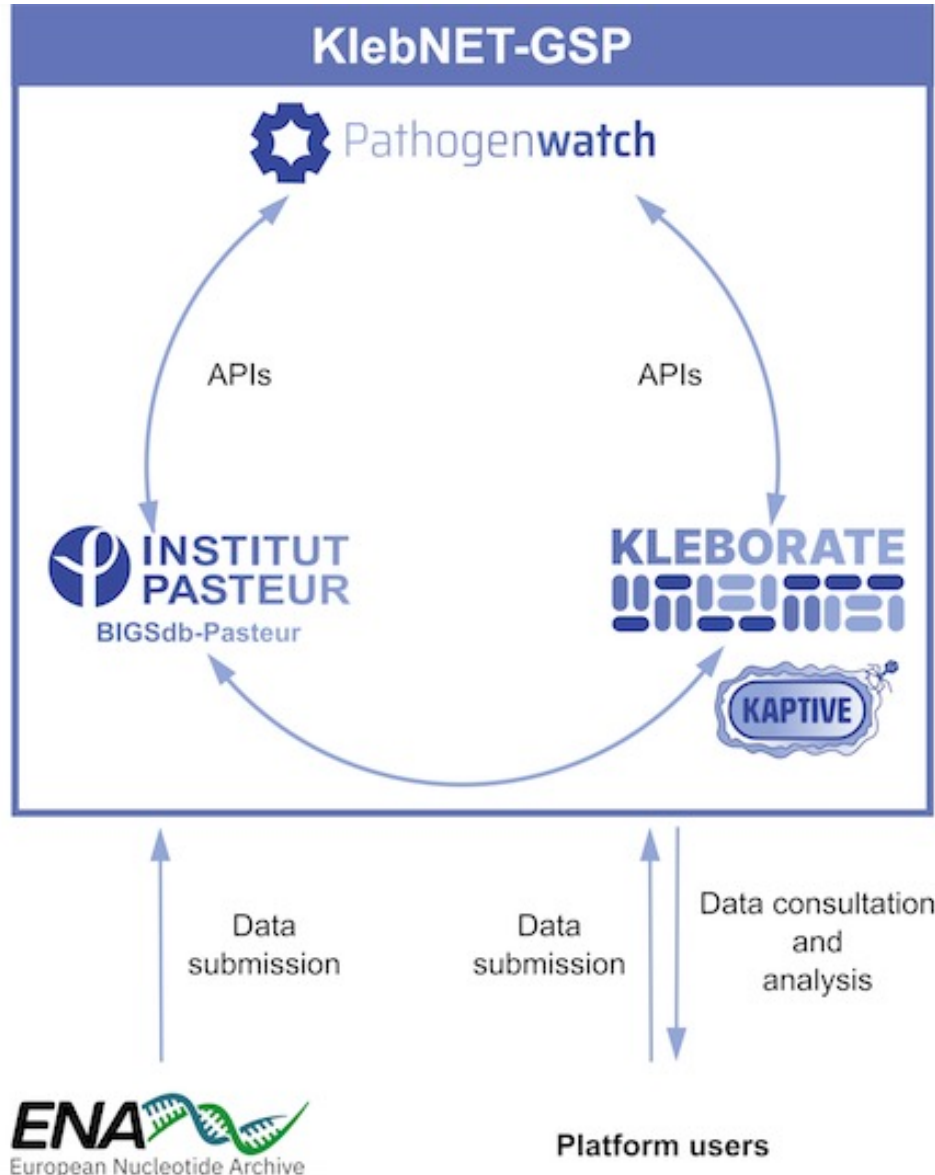
| MLST | | | | | | | | |
|------|-----|------|------|-----|------|-----|----|----------------|
| abcZ | adk | aroE | fumC | gdh | pdhC | pgm | ST | clonal complex |
| 1 | 3 | 3 | 1 | 4 | 2 | 3 | 4 | ST-4 complex |

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

Genomic surveillance: stronger together



KlebNET-GSP



- KlebNET-Genomic Surveillance Platform (<https://klebnet.org>)

Networking

Microbial & Sequencing
capacity

Analytical tools &
Databasing

Reporting
(clinical, public health)



Kat Holt (LSHTM)
David Aanensen (Oxford U)
Sylvain Brisse (Institut Pasteur)
& teams



Synchronization of ENA,
PW and BIGSdb-Pasteur

Interoperability

<https://bigsdb.pasteur.fr/>



<https://pathogen.watch/>



<http://www.genomicepidemiology.org/>



Center for Genomic Epidemiology

Nomenclature service

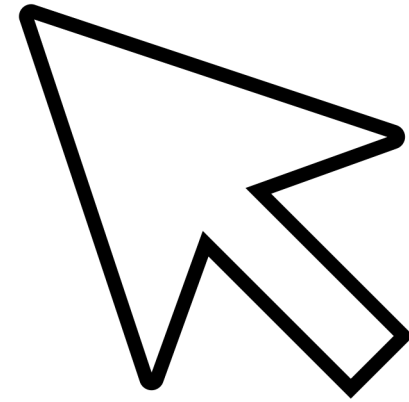
Collections

Taxonomic classification

Broad typing analysis

FIGURES

TRAVAUX --- PRATIQUES



Klebsiella pneumoniae PasteurMLST database



Record versions: current; Record creation: all time

Isolate count

11,690

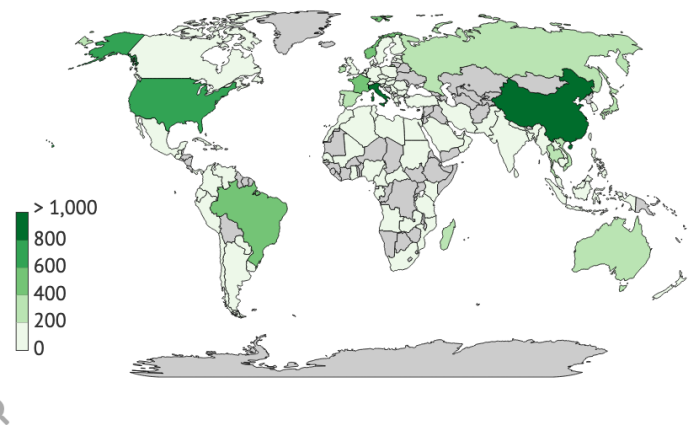
▲ 2 [month]

Genome count

6,433

▲ 2 [month]

Country

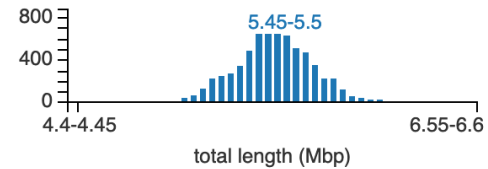


Continent

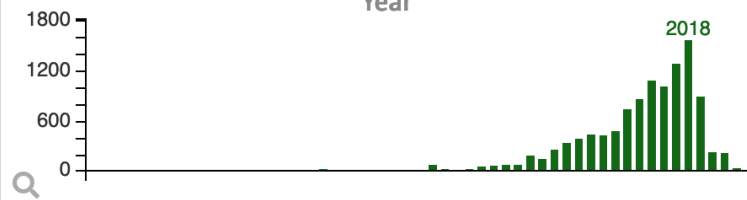
Top 5 values

| Value | Frequency |
|---------------|-----------|
| Europe | 5,058 |
| Asia | 3,662 |
| North America | 934 |
| Africa | 923 |
| South America | 573 |

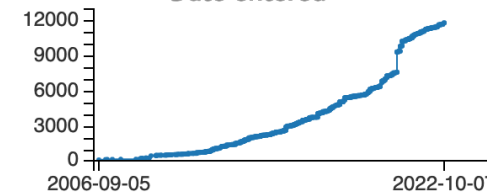
Sequence size



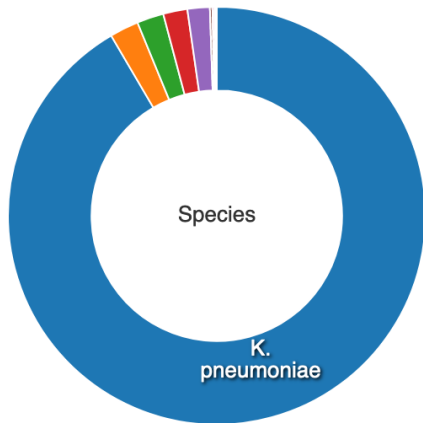
Year



Date entered



Species



LOG IN

SEARCH



SUBMISSIONS

PROJECTS



EXPORT



ANALYSIS



BLAST
BURST
Codon usage
Field breakdown
Gene presence
Genome comparator
GrapeTree
In silico PCR
iTOL
Microreact
PhyloViz
Publications
Sequence bin breakdown

<https://bigbdb.pasteur.fr>

- Search profile ST1
- Which isolates have profile ST1?
- What ST has profile 1-1-1-1-1-1-1 ?

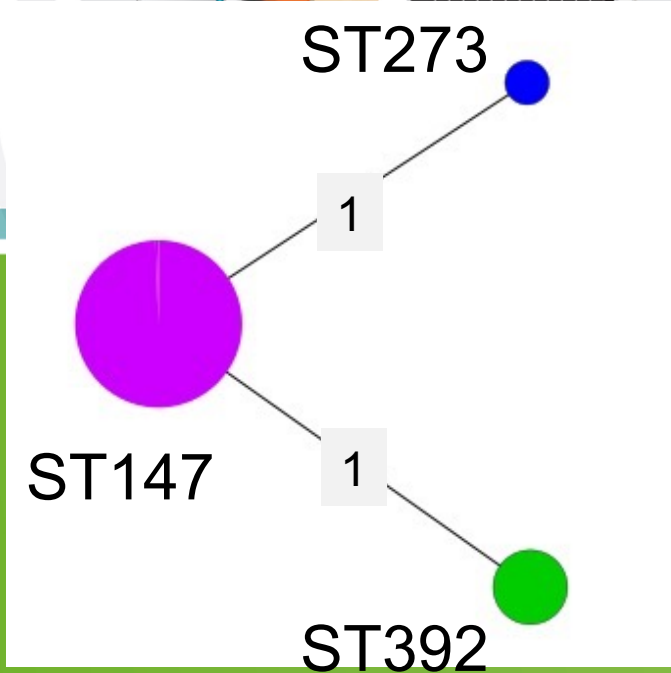
Klebsiella pneumoniae clonal group 147

MStree based on **MLST (7 genes)**

(Diancourt et al., J Clin Microbiol 2005)

MStree based on **cgMLST (629 genes)**

(Hennart et al. Mol Biol Evol 2022)



FIGURES

MStree: minimum spanning tree

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/primary metadata fields


id = Enter value... + i

Filters

Publication: Select options i

Project: 1 selected i

9 records returned. Click the hyperlinks for detailed information.

| Isolate fields  | | | | | | | | | | |
|--|------------|---------|-----------------------|--------------------|--------------|-------------|-------|----------------|-----------|-------------------|
| id | isolate | aliases | taxonomic designation | phylogroup from ST | world region | source type | host | source details | infection | other source info |
| 7091 | NMI1203_15 | | K. pneumoniae | Kp1 (7 loci) | Europe | | human | | | |
| 7101 | NMI1799_15 | | K. pneumoniae | Kp1 (7 loci) | Europe | | human | | | |
| 7106 | NMI2273_16 | | K. pneumoniae | Kp1 (7 loci) | Europe | | human | | | |
| 14499 | NMI650_19 | | K. pneumoniae | | Europe | Human | Human | Unknown | | |

Analysis tools

Breakdown: Fields Two Field Combinations Polymorphic sites Publications **Sequence bin**

Analysis: BURST Codons Gene Presence Genome Comparator BLAST

Export: Dataset **Contigs** Sequences

Third party: **GrapeTree** iTOL PhyloViz Microreact

GrapeTree :

- Step 1: search BIGSdb-Pasteur Klebsiella database

- Step 2:
1. MLST

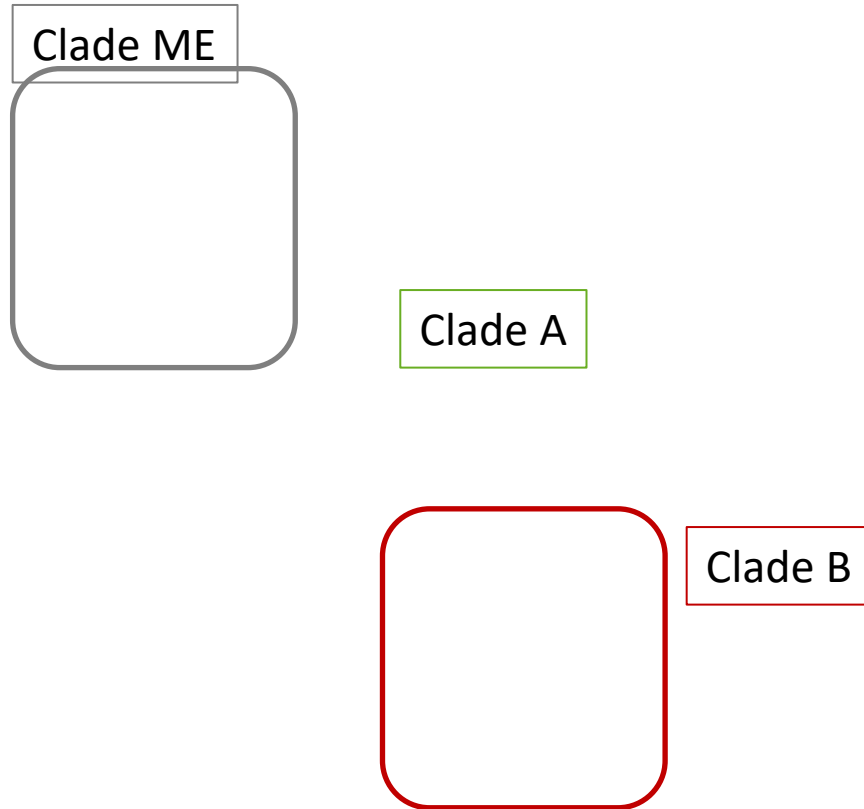
2. scgMLST_S: 629 loci

- Step 3: GrapeTree

<https://bigsddb.pasteur.fr>

- Search all strains from your own country
- Are there strains from food from your country?
- What is the less frequent Klebsiella ST in your country?
- Do you find it elsewhere in the world?
- How are your country strains related to other countries' isolates?

Are MStrees based on allele profiles, phylogenies?



FIGURES

Martin *et al.*, 2022 PNAS

- Minimum spanning tree: hypothesizes that ancestral nodes are sampled
- OK to visualize short-term relationships
- Unrooted -> no directionality of evolution

Data exploration using MStrees

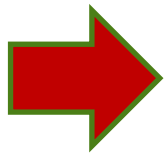


Data: Martin et al., 2022 PNAS: Anatomy of an extensively drug-resistant *Klebsiella pneumoniae* outbreak in Tuscany, Italy (BIGSdb-Pasteur public project 77)

FIGURES

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

Acknowledgements

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