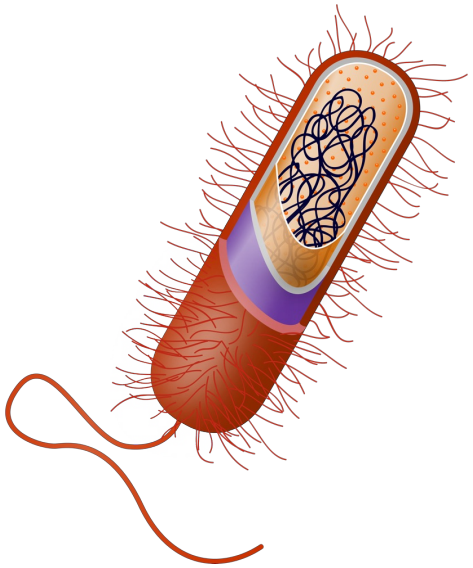
Introduction to Typing

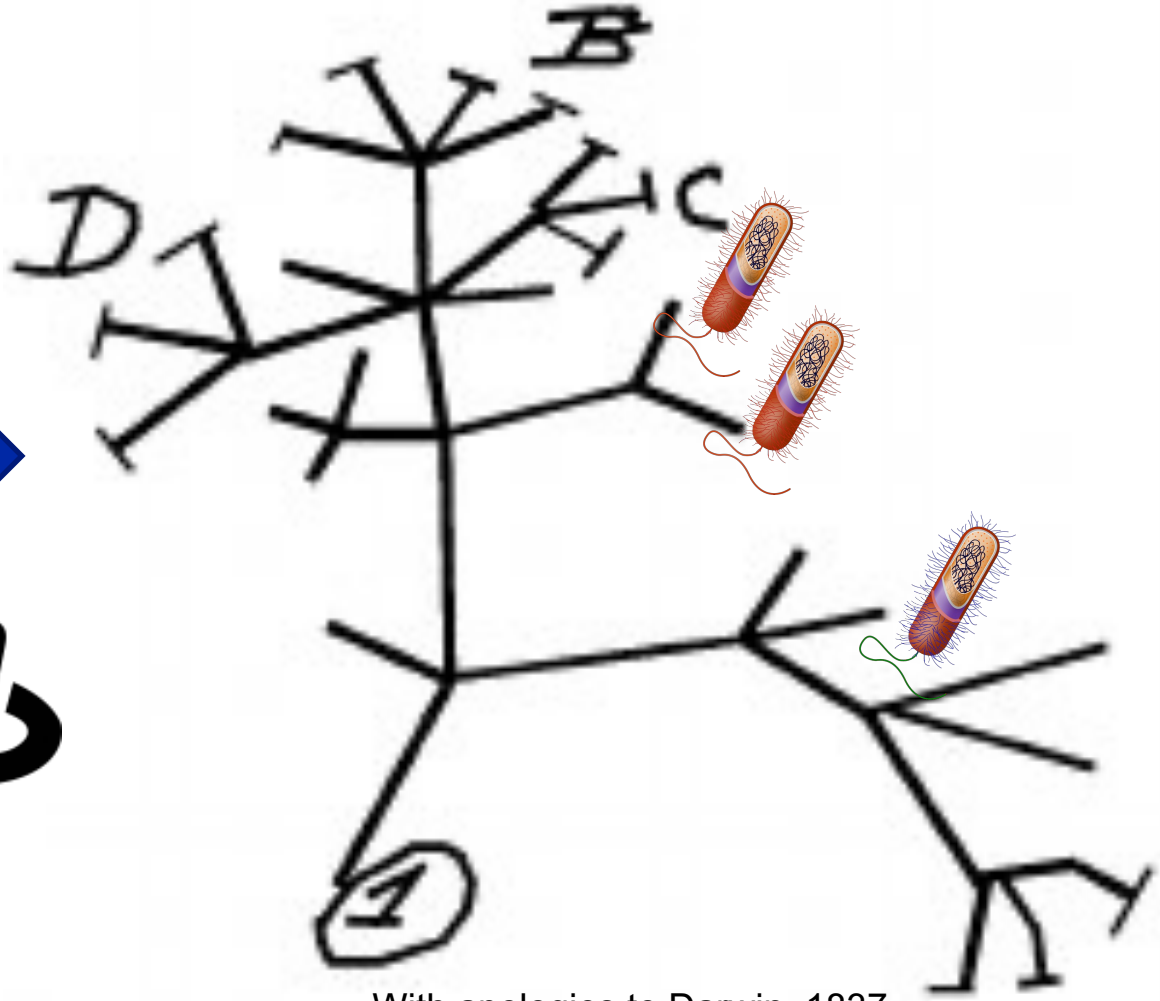
Helena Seth-Smith PhD

23.03.2023

Table of Contents

What is Typing?
A method of categorizing microbes





With apologies to Darwin, 1837

Conventional methods of typing

"Microbial typing is often employed to determine the source and routes of infections, confirm or rule out outbreaks, trace cross-transmission of healthcare-associated pathogens, recognize virulent strains and evaluate the effectiveness of control measures" (Ranjbar *New Microbiologica* 2014)

Conventional typing was phenotypic:

- Gram staining
- Serotyping: based on surface characteristics recognized by immune system (eg rabbit antisera)
- Phage typing

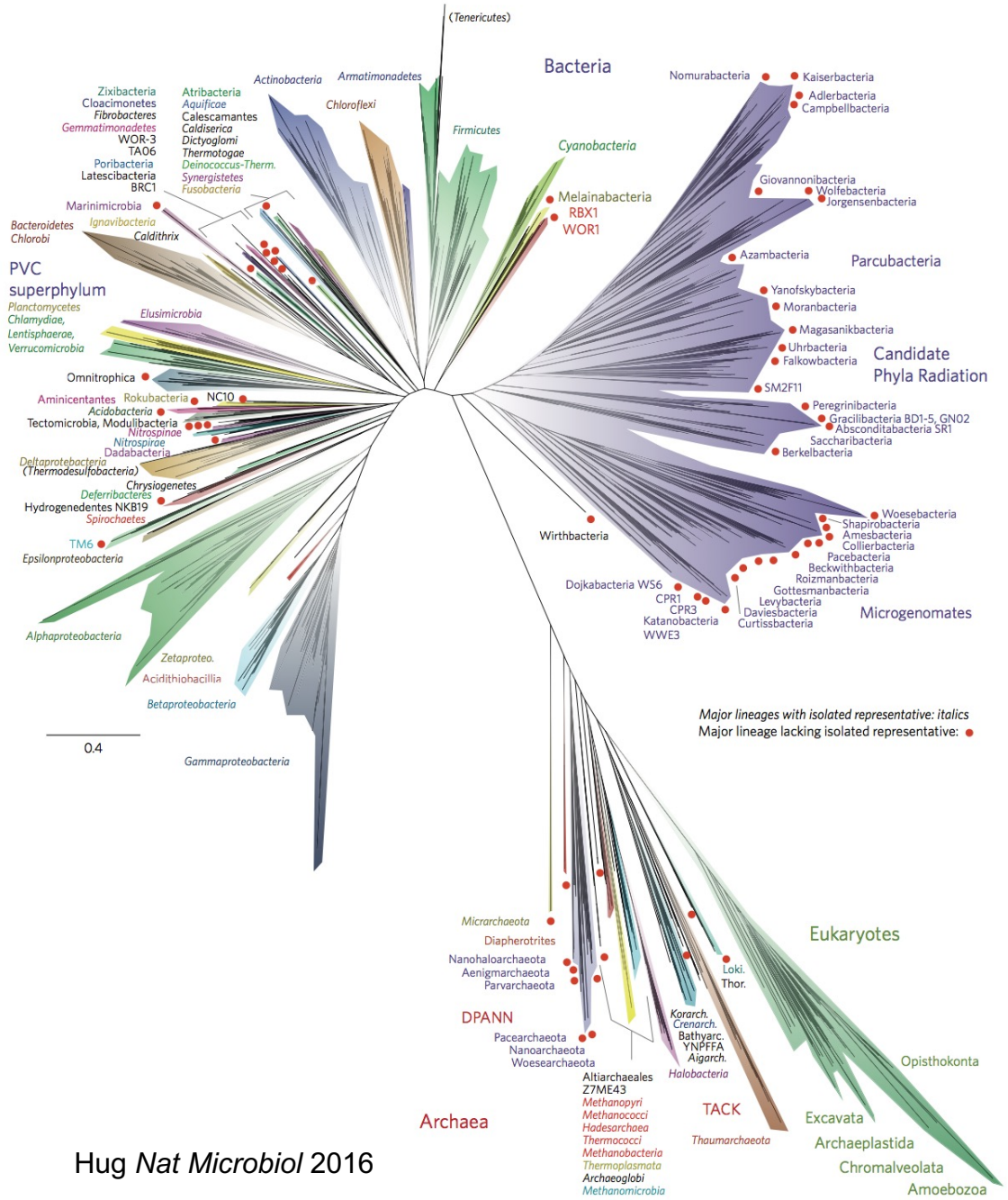
Generally considered too

variable, labour intensive, time-consuming, insufficiently discriminatory and poorly reproducible

Typing then moved into "molecular" or DNA-based era

Typing: why molecular?

- Genomes **define** the ancestry and relatedness of isolates
- Any way to access this information can provide better typing results



Hug *Nat Microbiol* 2016

Molecular typing

Analyses discriminatory parts of the genome in the lab

– often amplification of specific regions / loci followed by separation on a gel and reading the pattern

Pulsed Field Gel Electrophoresis (PFGE)

Amplified fragment length polymorphisms (AFLP)

Random amplification of polymorphic DNA (RAPD)

Repetitive-element polymerase chain reaction (rep-PCR)

Variable-number tandem repeat (VNTR)

Staphylococcus aureus protein A gene-typing (spa)

Multilocus Enzyme Electrophoresis (MLEE – protein variant based)

Ribotyping (*Clostridioides difficile*)

Multi-locus sequence typing (MLST)

Different approaches are more suitable for different species

Molecular typing

Some techniques useful across the whole species

– And some offer higher resolution within closely related strains (“clones”)

Pulsed Field Gel Electrophoresis (PFGE)

Amplified fragment length polymorphisms (AFLP)

Random amplification of polymorphic DNA (RAPD)

Repetitive-element polymerase chain reaction (rep-PCR)

Variable-number tandem repeat (VNTR)

Staphylococcus aureus protein A gene-typing (spa)

Multilocus Enzyme Electrophoresis (MLEE – protein variant based)

Ribotyping (*Clostridioides difficile*)

Multi-locus sequence typing (MLST)

Because they are based on different genomic features

Pulsed Field Gel Electrophoresis

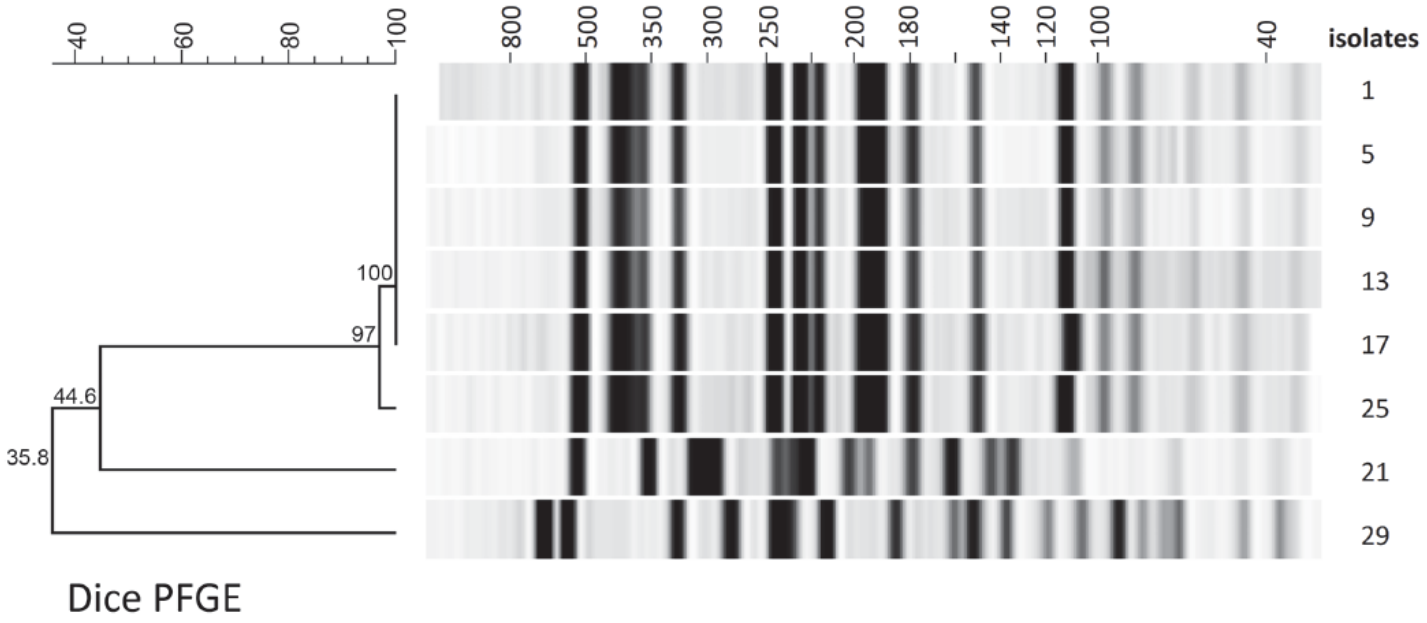
Was reference standard for many years

Comparison of band patterns

Algorithm determines the phylogeny

Few databases available

Low reproducibility between centers



E. coli PFGE
Outbreak and two unrelated isolates
Egli *PLoS ONE* 2015

Salmonella serotyping (and more)

- *Salmonella* Typhi (*Salmonella enterica* subspecies *enterica* serotype Typhi) is a serovar of a disease-syndrome defined group of a subspecies
- *Salmonella* serotyping depends on specific agglutination reactions with adsorbed antisera specific for O or H antigens
- Antigen + appropriate antibody = agglutination

Universe84a.com

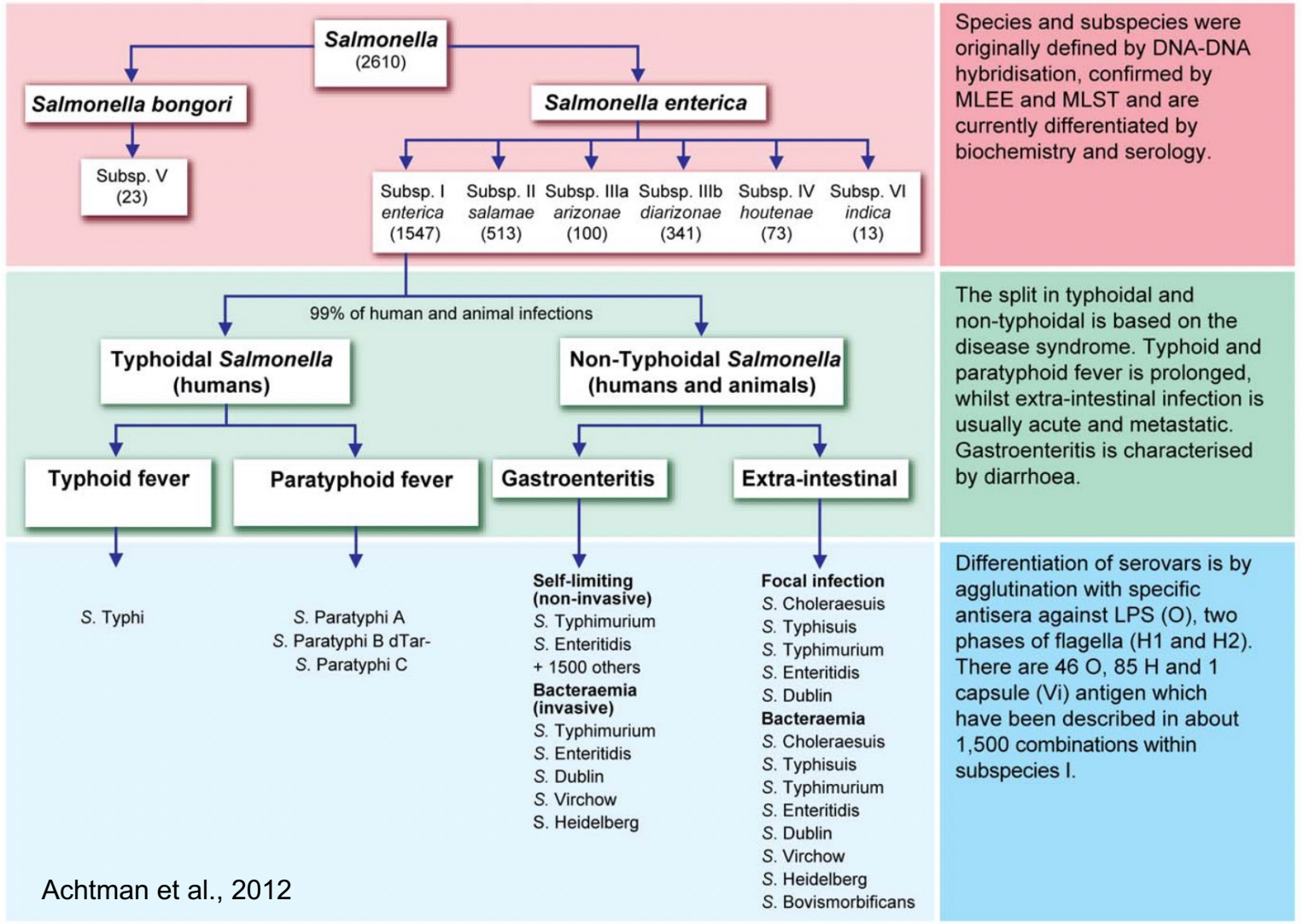
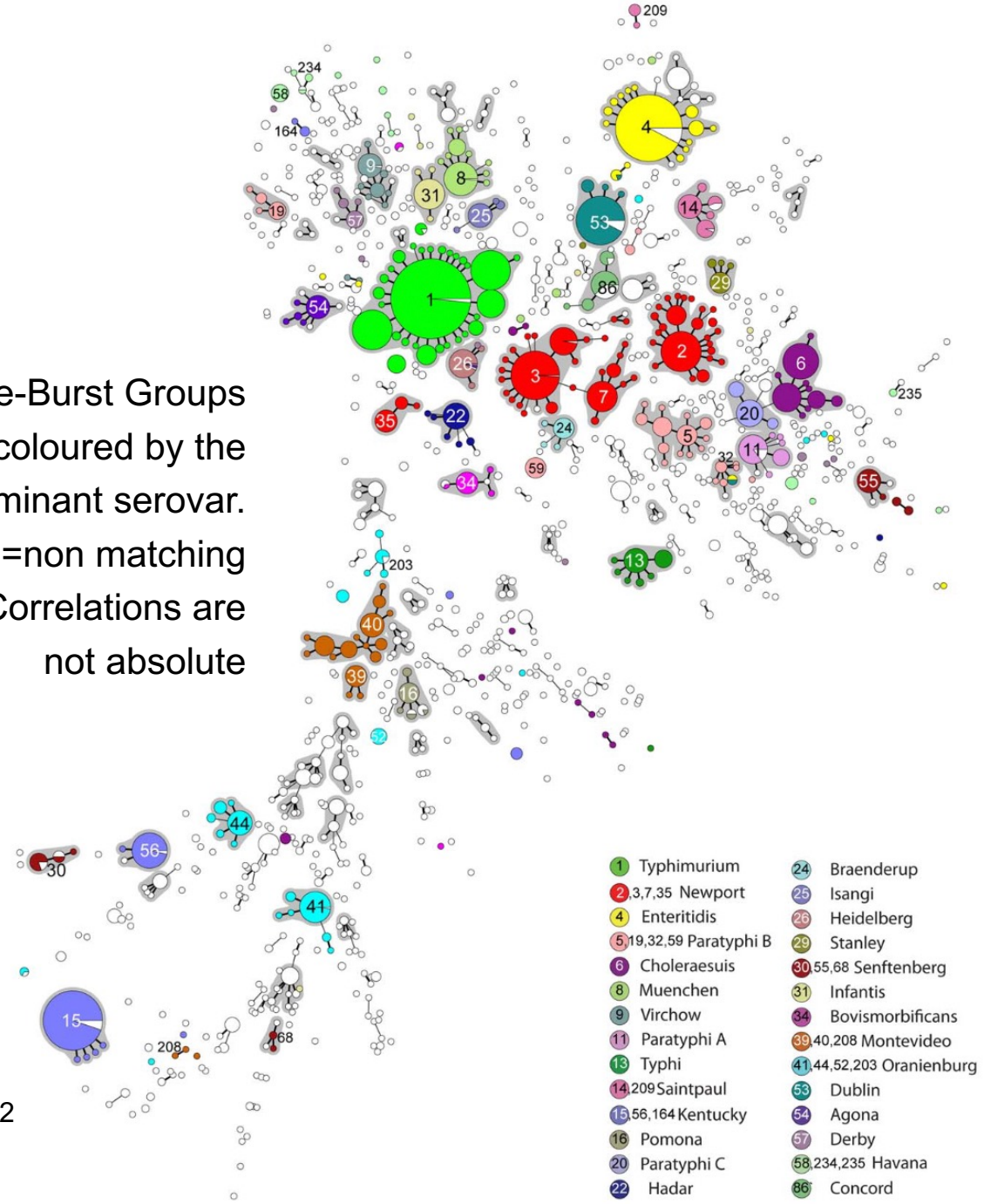


Figure 1. General overview of the current classification of *Salmonella enterica*.
doi:10.1371/journal.ppat.1002776.g001

Salmonella MLST

- *Salmonella* Typhi (*Salmonella enterica* subspecies *enterica* serotype Typhi) is a serovar of a disease-syndrome defined group of a subspecies
- *Salmonella* serotyping depends on specific agglutination reactions with adsorbed antisera specific for O or H antigens
- Comparison with MLST (7 gene fragments): “serovar designations confounded genetically unrelated isolates and failed to recognize natural evolutionary groupings”

Groups / e-Burst Groups are coloured by the dominant serovar. White=non matching serovar. Correlations are not absolute

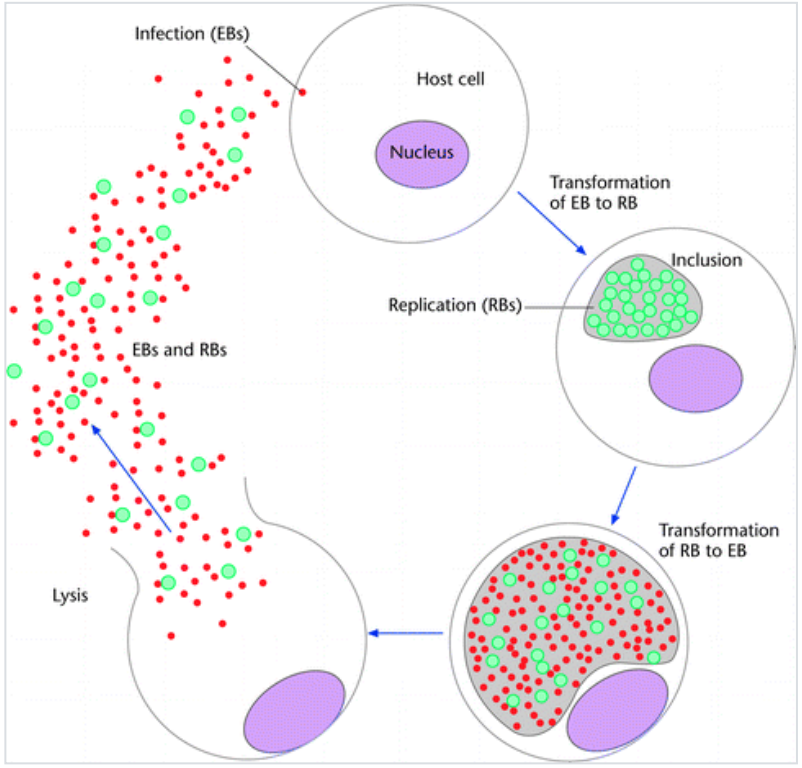


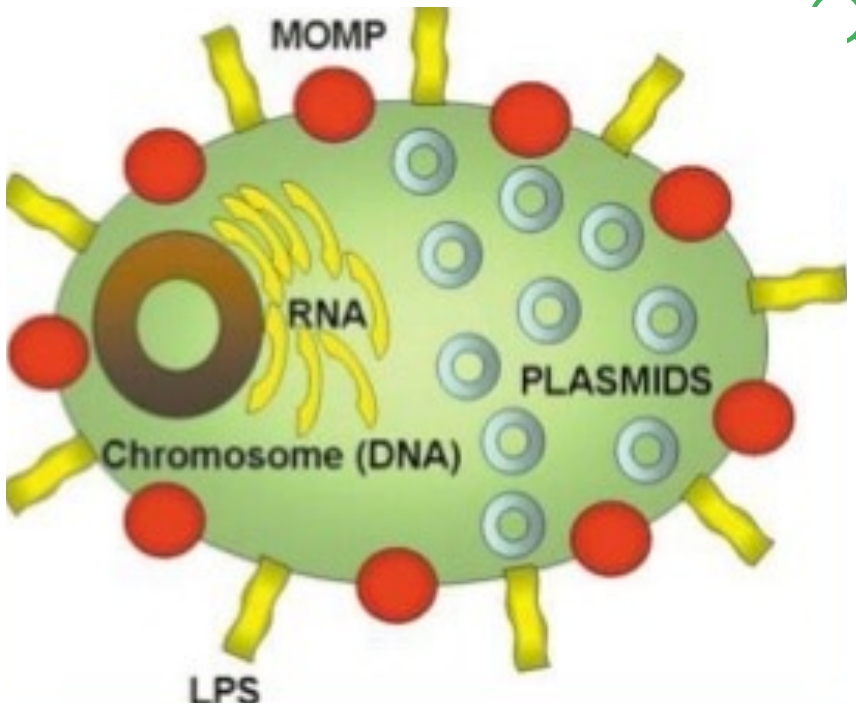
Achtman et al., 2012

Figure 2. Minimal spanning tree (MSTree) of MLST data on 4257 isolates of *S. enterica* subspecies *enterica*. Each circle corresponds to

C. trachomatis: serotyping became genotyping

- Gram negative
- Intracellular
- Small genome ~1 Mb
- Few gene differences across species

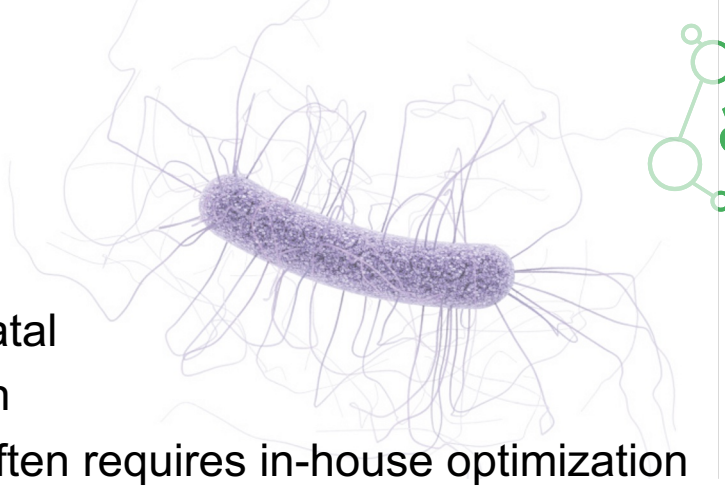




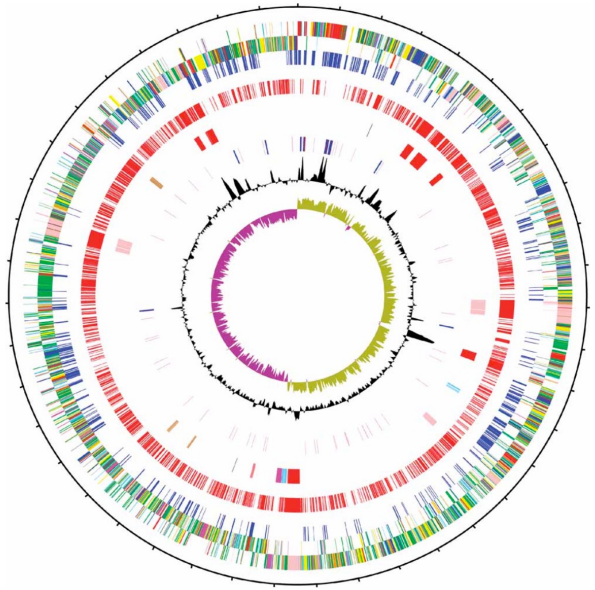
- MOMP: 60% of outer membrane protein mass
  - serotyping with monoclonal antibodies from eg. HeLa 229 cells
- MOMP encoded by gene *ompA* -> genotyping

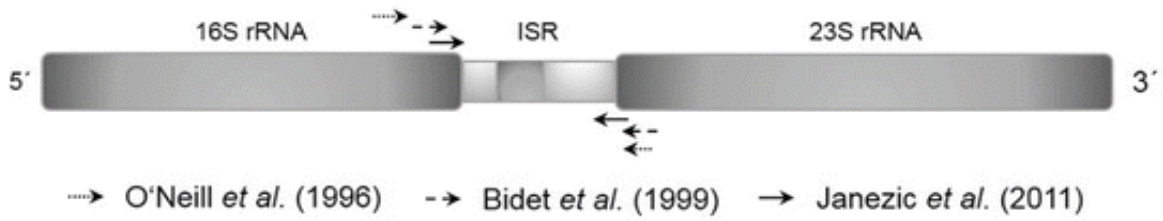
Stephens *Science* 1998 (Trachoma); Carlson *Inl* 2005 (STI / Urogenital); Thomson *Genome Res* 2008 (LGV); Seth-Smith *BMC Genomics* 2009

Clostridioides difficile: ribotyping



- Important nosocomial pathogen
- Asymptomatic -> *C. difficile* infection (CDI) -> severe colitis, sepsis, fatal
- PCR-ribotyping commonly used typing tool: good strain discrimination
- BUT not fully portable between laboratories, labour intensive, slow, often requires in-house optimization
- Hypervirulent lineages have been defined by ribotype: RT027 and RT078
- What is the connection between ribotype, virulence and genome?





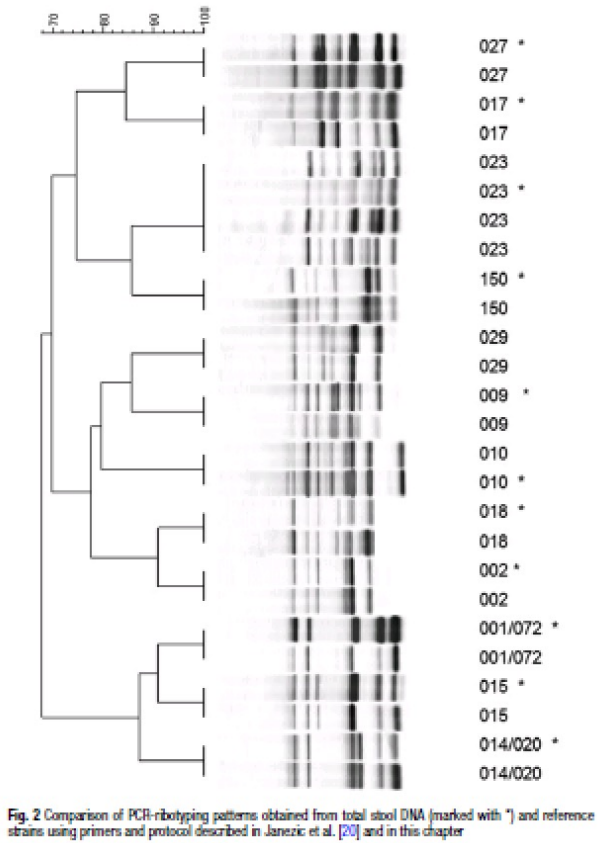
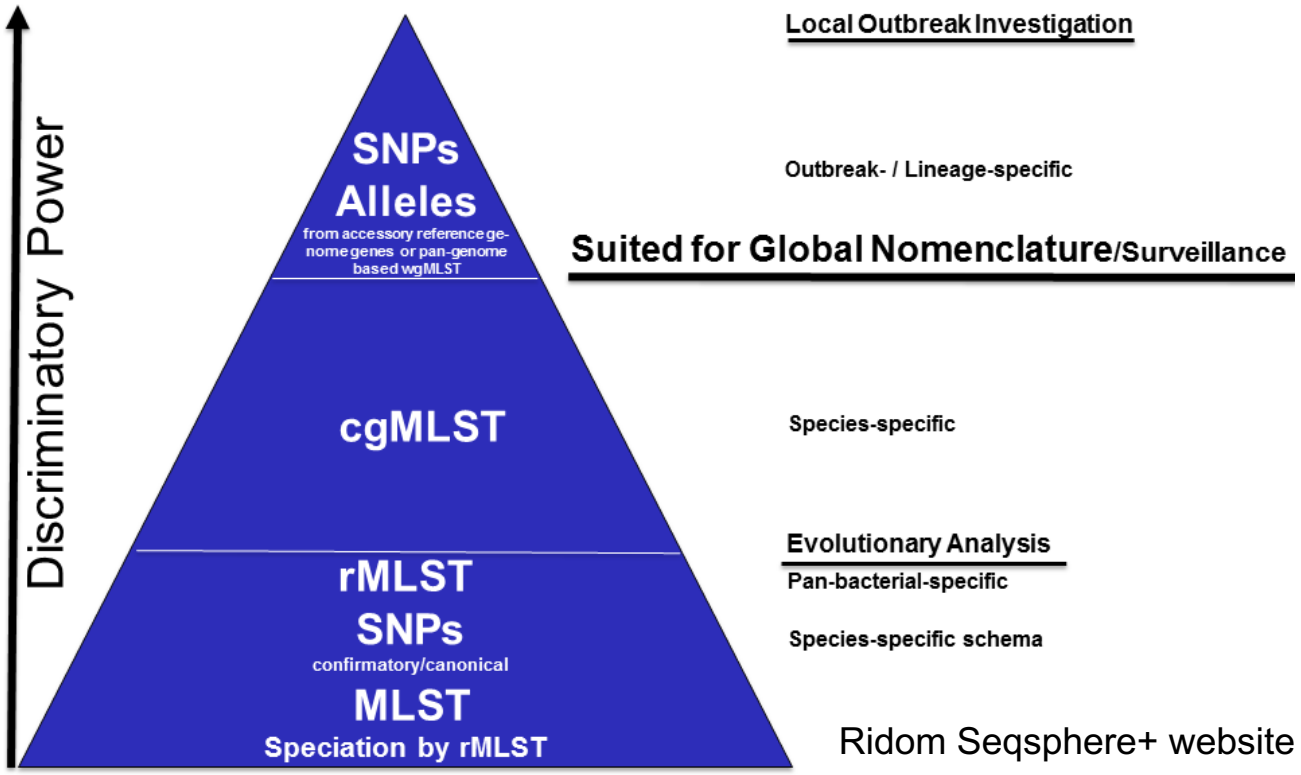


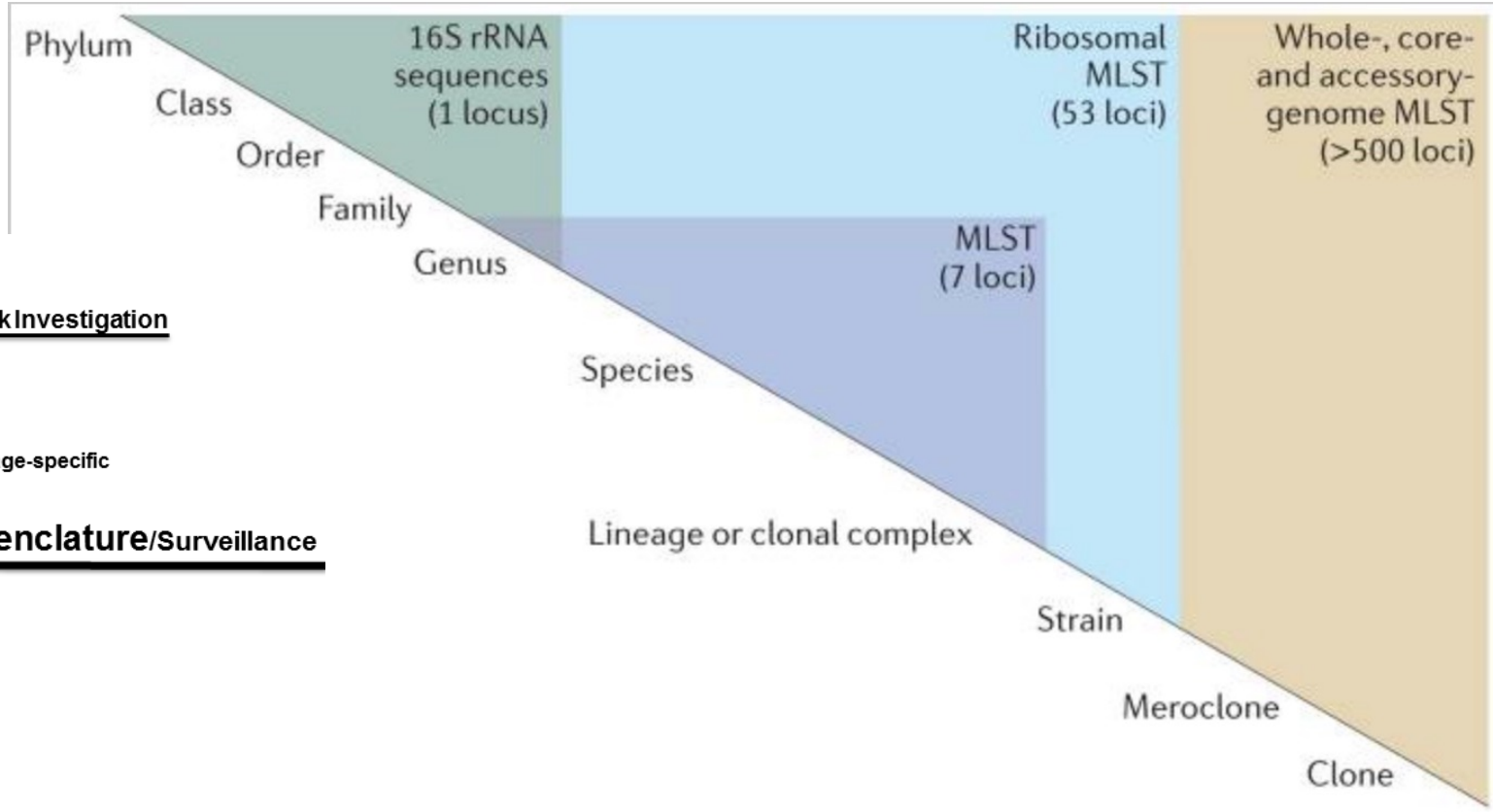
Fig. 2 Comparison of PCR-ribotyping patterns obtained from total stool DNA (marked with *) and reference strains using primers and protocol described in Janezic et al. [20] and in this chapter

Sebaihia *Nat Gen* 2006; Janezic *Clostridium difficile: Methods and Protocols* 2016; *C. difficile* Methods and Protocols, Springer

Genome-based typing methods: resolution



Ridom Seqsphere+ website

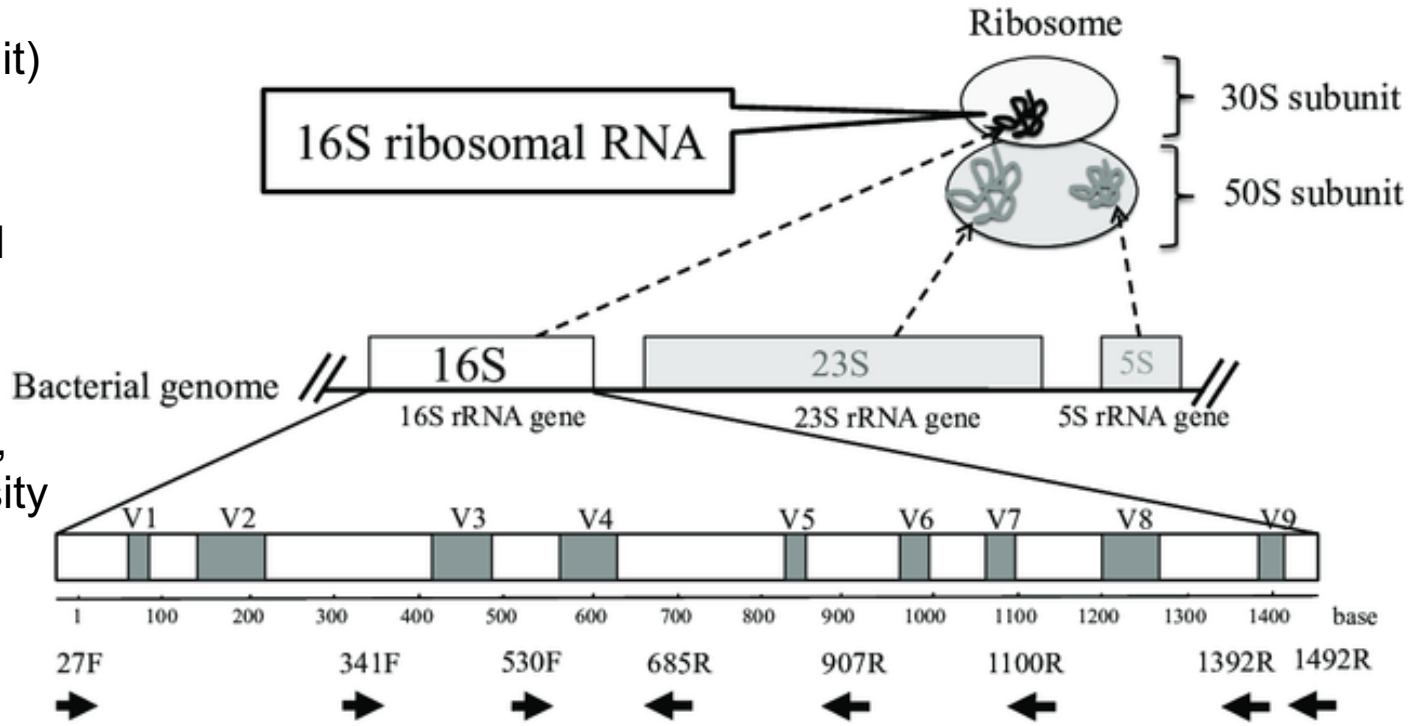


Maiden *Nat Rev Microbiol* 2013

16S ribosomal RNA typing

- 16S rRNA gene produces 16S rRNA
- Forms part of ribosome (part of 30S subunit)
- Sufficiently conserved / diverse to define species when sequenced
- V / variable regions are frequently targeted

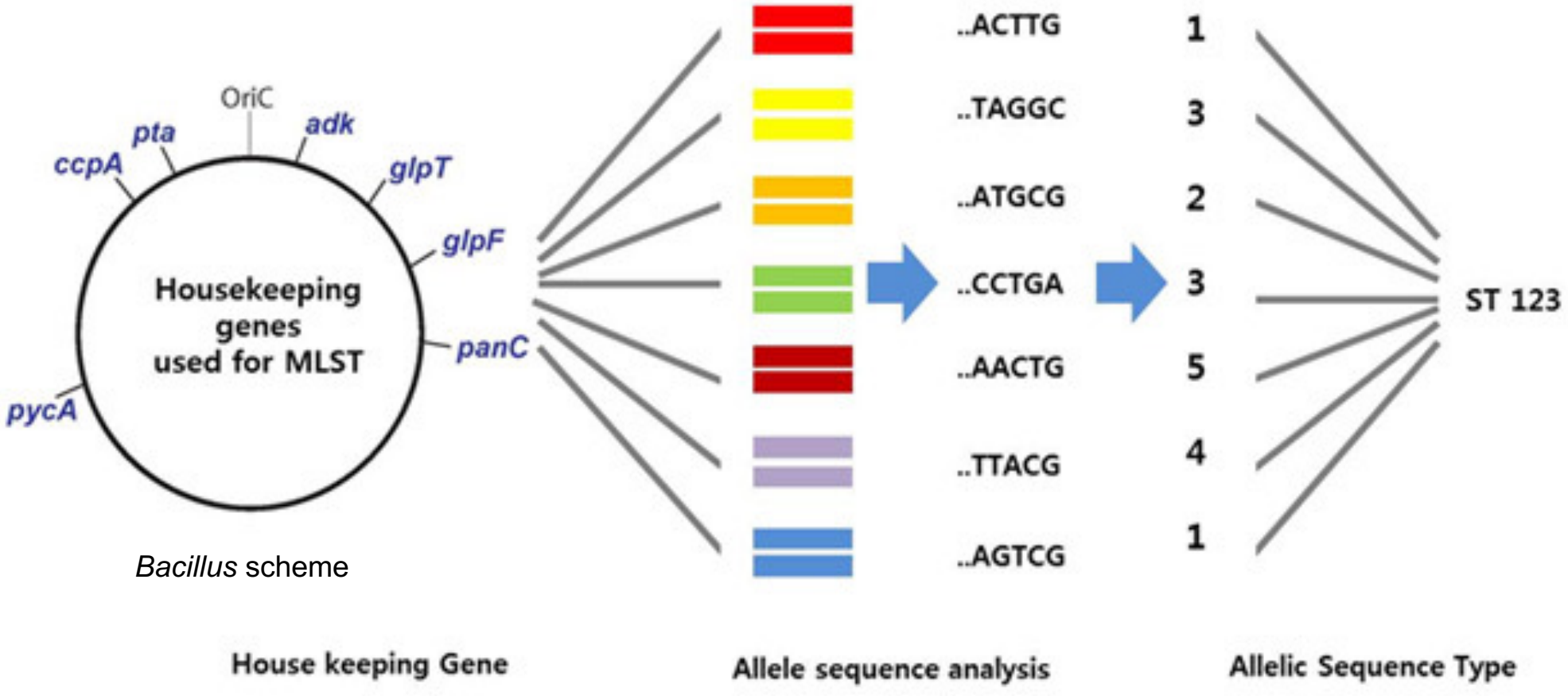
- Species correlation not always perfect
- Some genomes have several copies: 2-14, average 6, and they can have some diversity



DOI: 10.7888/juoeh.38.223

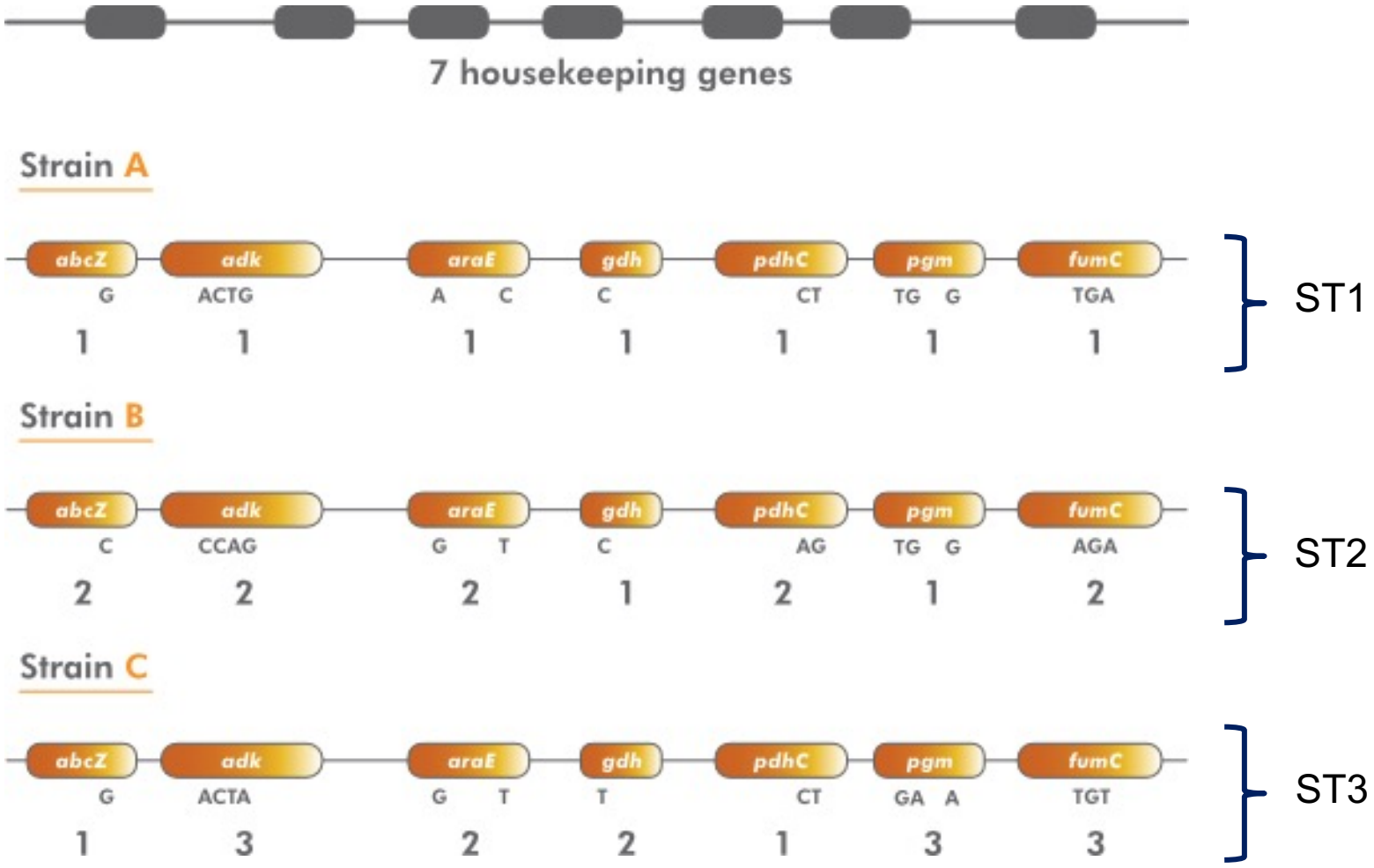
Typing methods: MLST

- Manageable number (5-7) of conserved (housekeeping) genes to represent genome diversity
- Originally by PCR and Sanger sequencing: data can also be extracted from WGS
- First in Maiden *PNAS* 1998; *Neisseria meningitidis*; 6 loci chosen to reflect resolution of MLEE



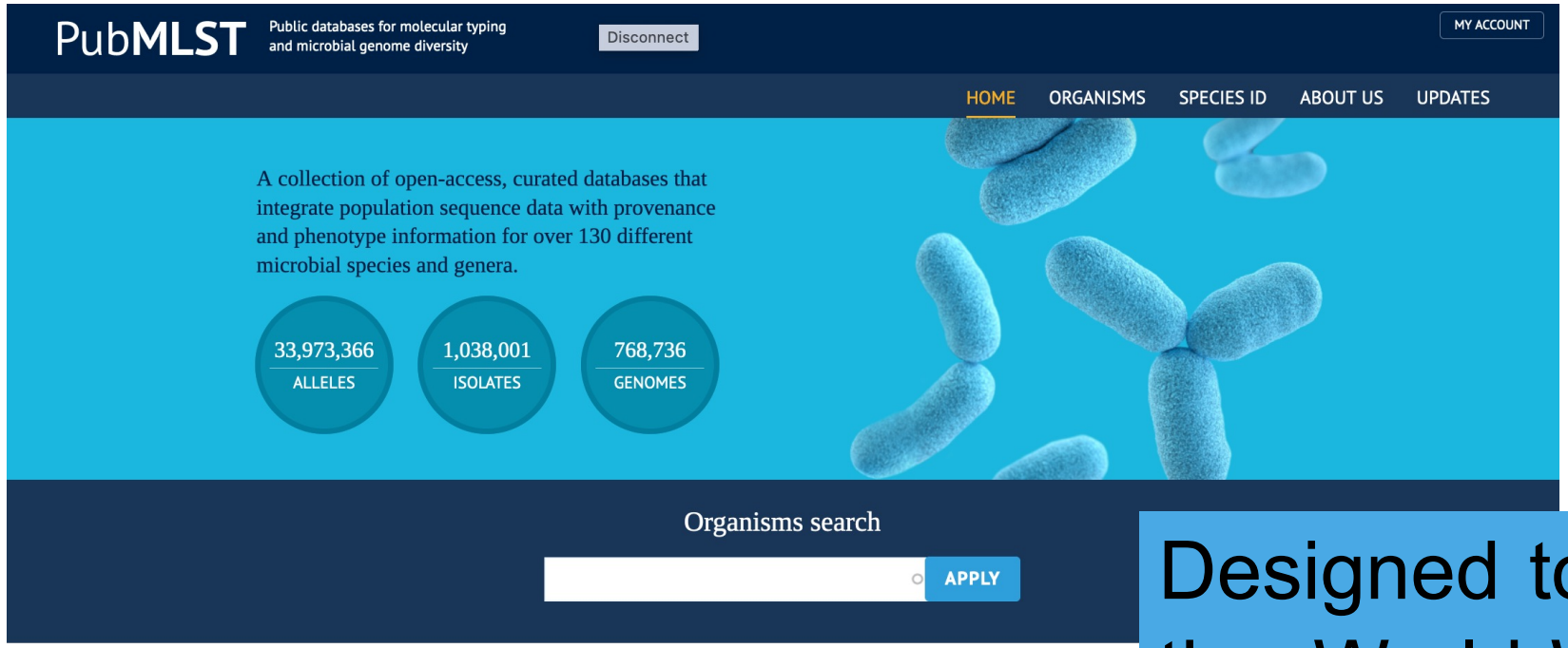
Bacillus scheme

How does MLST work?



Biomérieux, Applied Maths, Bionumerics website

PubMLST



- Schemes defined for many species: international, sharable…

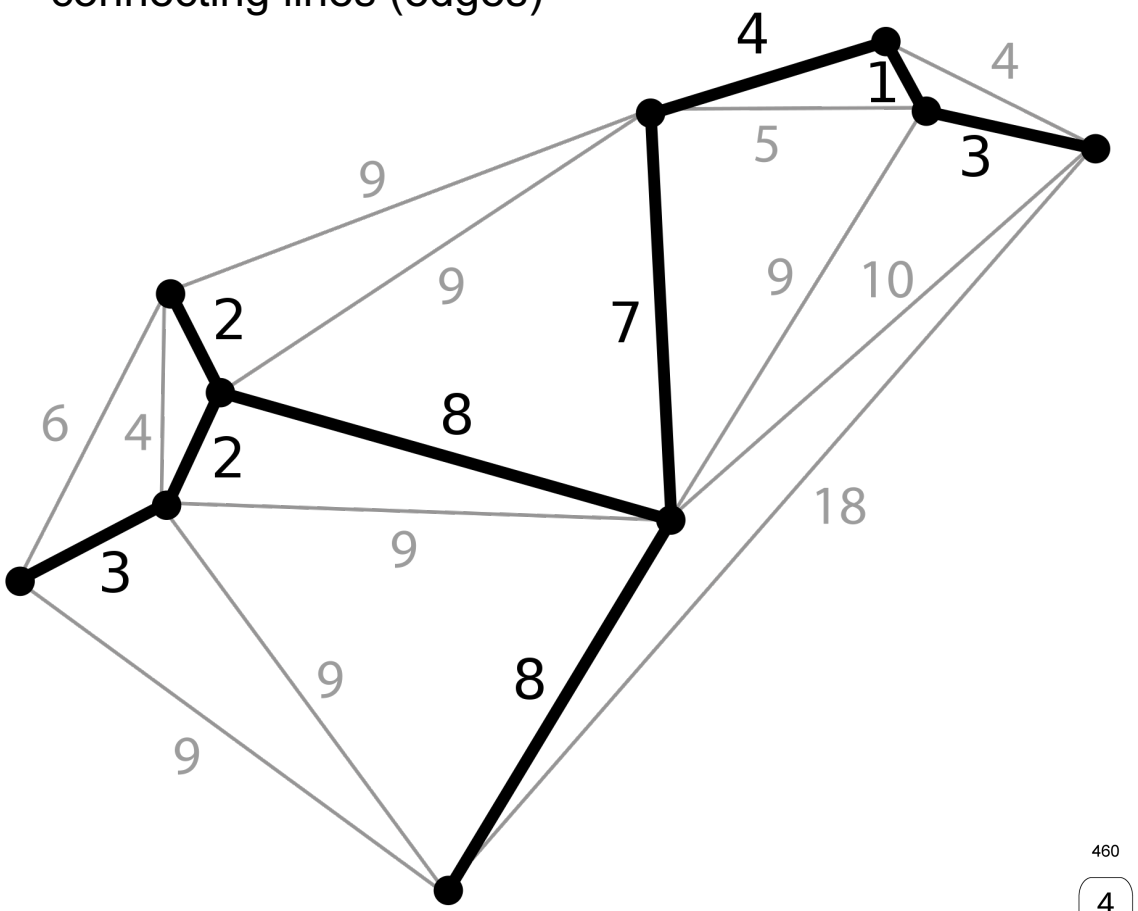
Designed to be portable “across the World-Wide Web site, thus enabling exchange of molecular typing data for global epidemiology via the Internet.”

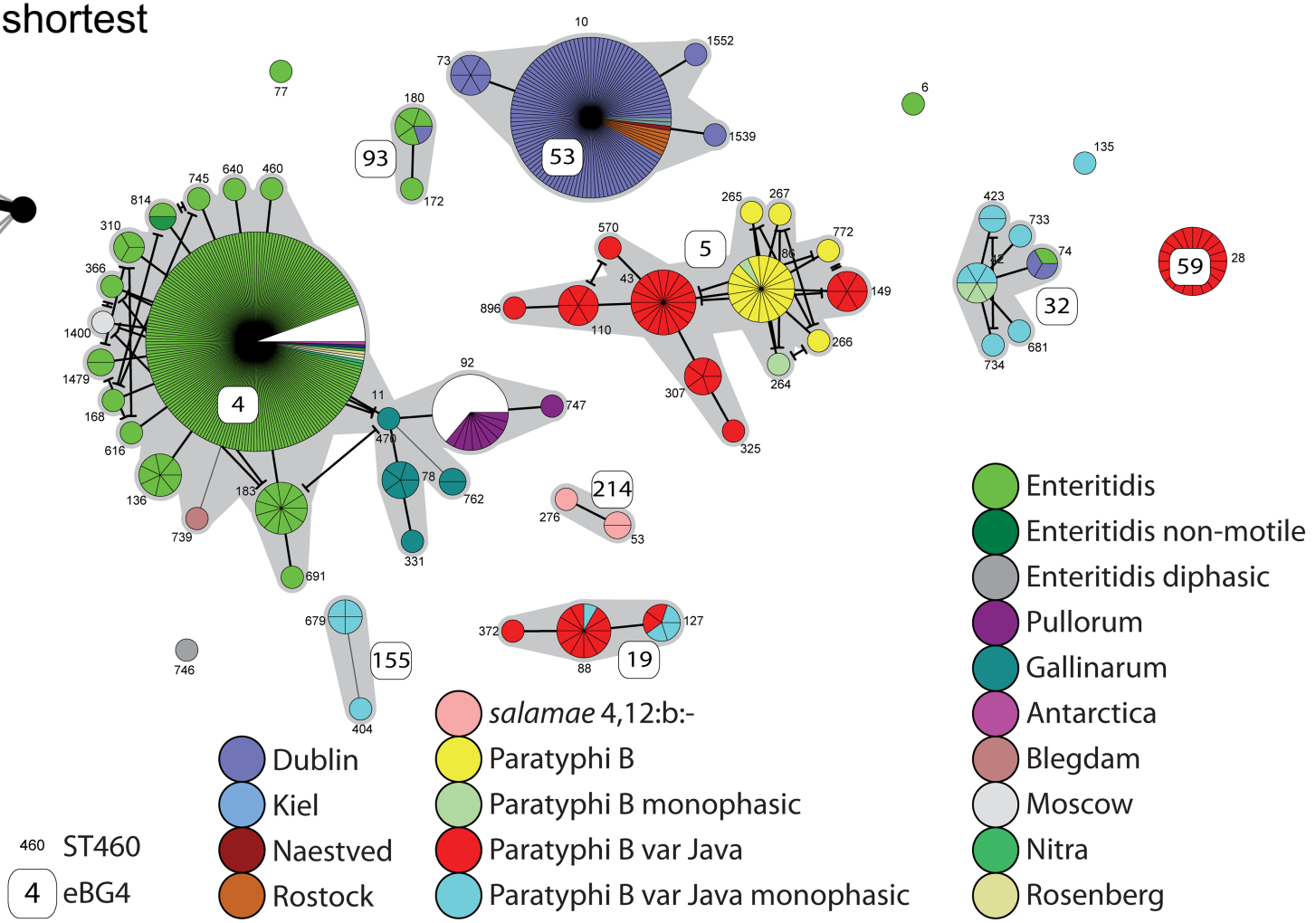
Maiden *PNAS* 1998

MLST representation: Minimum Spanning Tree

Achtman *PLoS Pathogens* 2012;
Salmonella enterica MLST vs serovar

A tree connecting all data points (nodes) using the shortest connecting lines (edges)



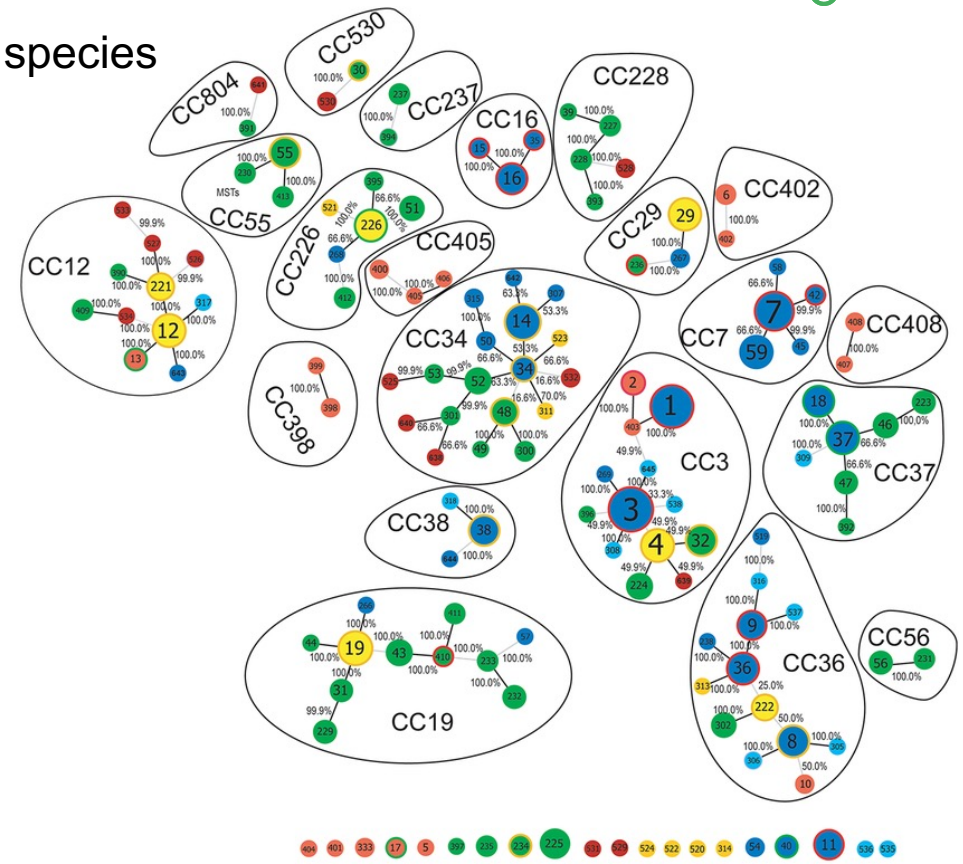


Clonal Complexes

Clonal complexes are groupings of MLST sequence types, in some species

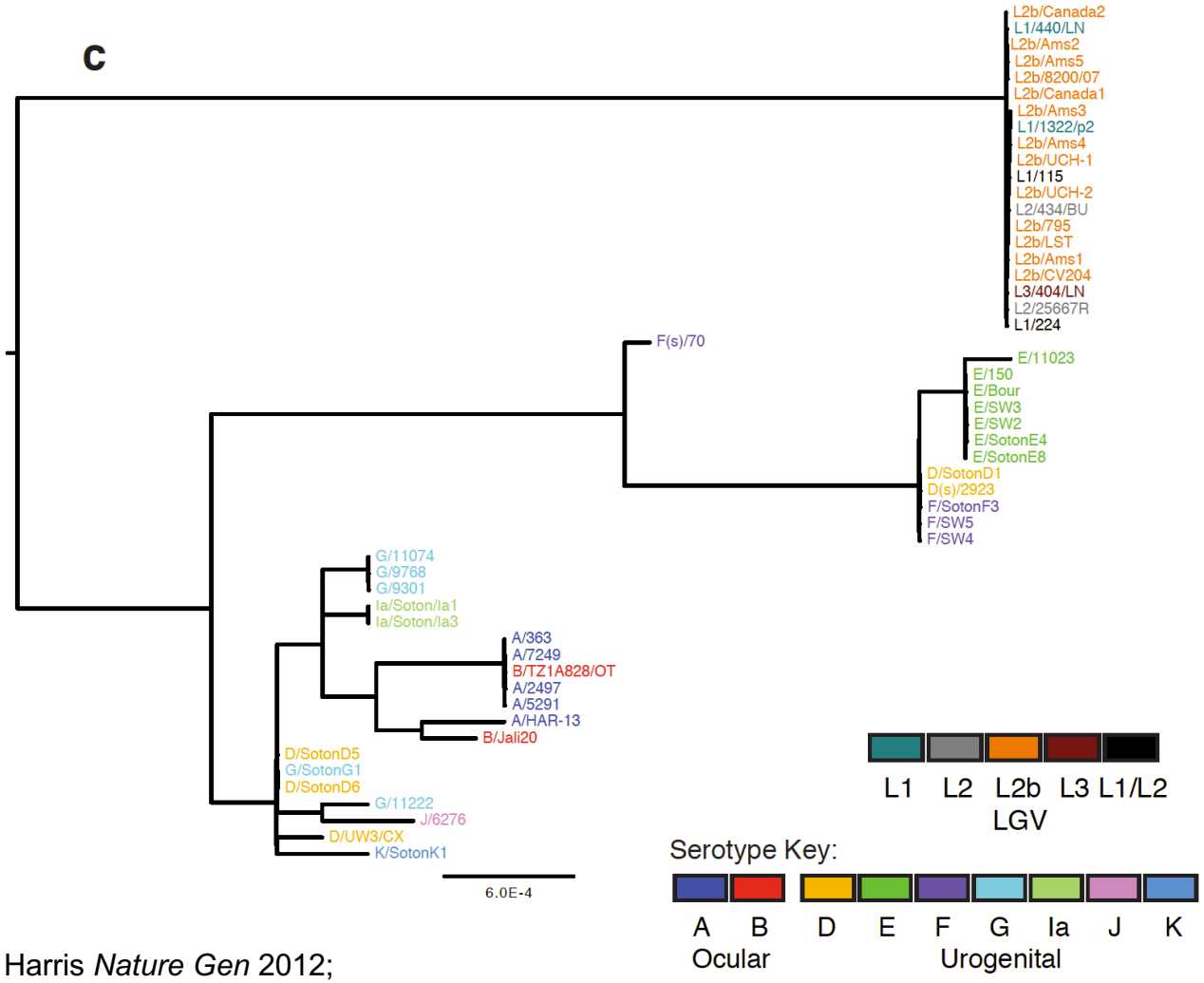
Grouped according to a specific number of shared alleles, eg:

- *Staphylococcus aureus:* 4+ shared loci
- *Borrelia burgdorferi*: single / double locus variants (figure)



10.1371/journal.pone.0149345

MLST representation:
SNP tree of concatenated loci

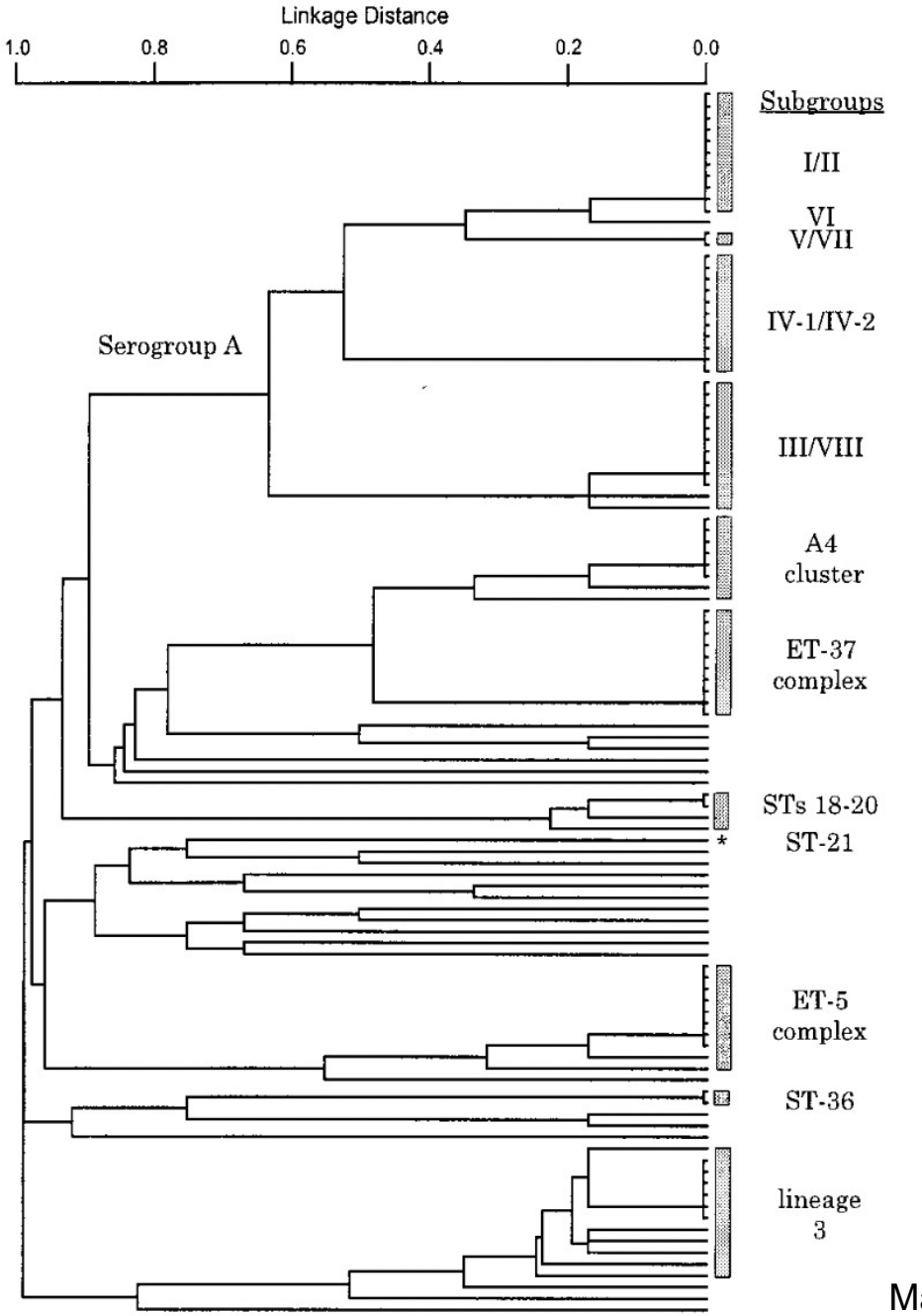


Harris *Nature Gen* 2012;
Chlamydia trachomatis

Neisseria meningitidis: hypervirulent lineages and MLST

- Most important hyper-virulent lineages are subgroups I, III, and IV-Other, by MLEE

- Comparison of MLST results:
  - serogroup A strains formed a distinct cluster of lineages
  - major subgroups associated with epidemic meningitis (I, III, and IV-1) were easily distinguished

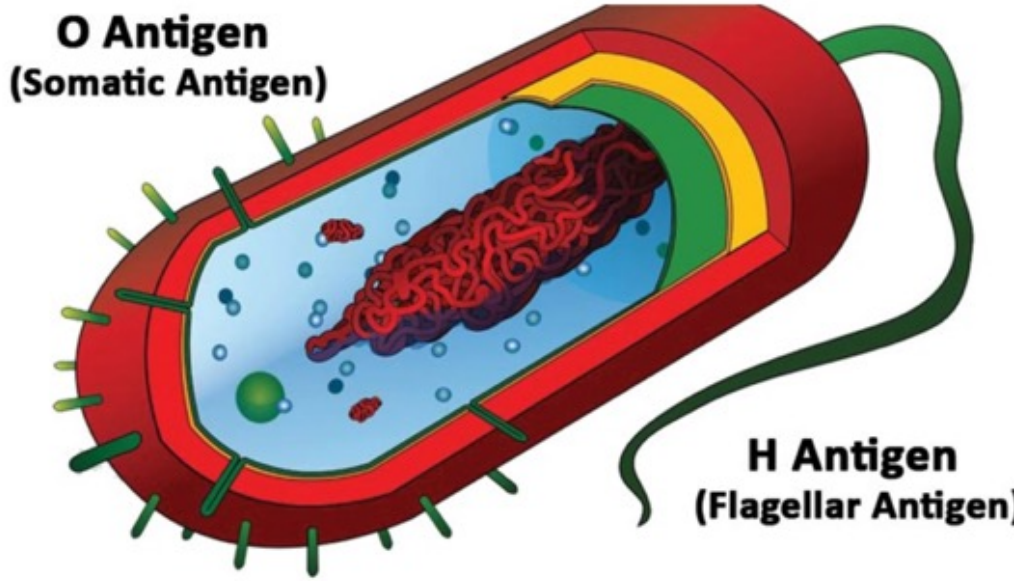


Maiden *PNAS* 1998

FIG. 2. Dendrogram of genetic relationships among 107 strains

E. coli typing – disease association

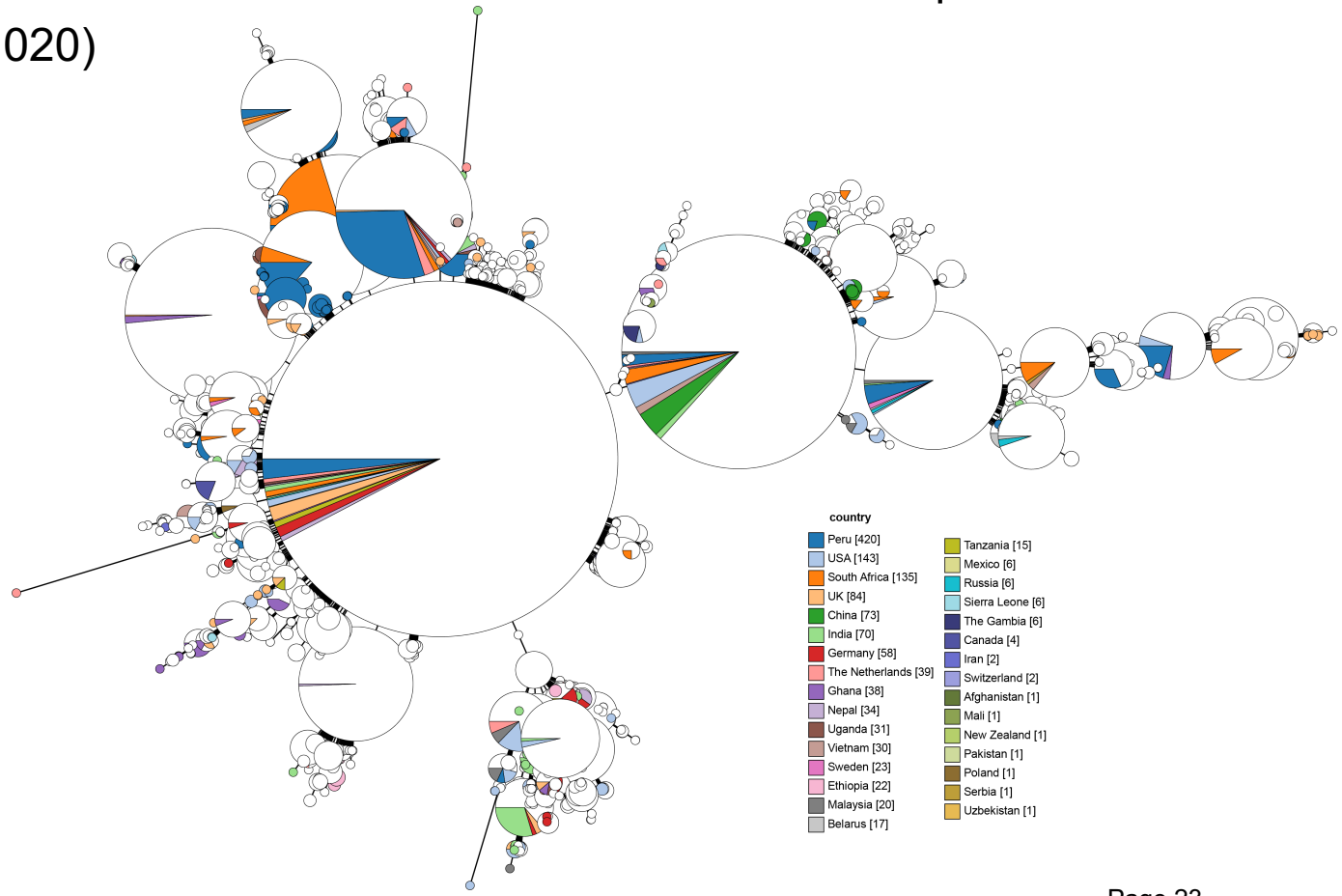
- *E. coli* O157:H7 commonly associated with disease in humans
  - Based on O and H antigen serotyping
  - Also Shiga toxin typing based on serology / amino acid sequence predicted from genome
- Different molecular techniques give different resolutions:
  - Randomly amplified polymorphic DNA (RAPD) and pulsed-field gel electrophoresis (PFGE) given resolution within O157:H7. Multi-locus sequence typing (MLST) doesn't. Typing using techniques using whole genome data provide highest resolution. (doi.org/10.1038/s41538-021-00097-0)



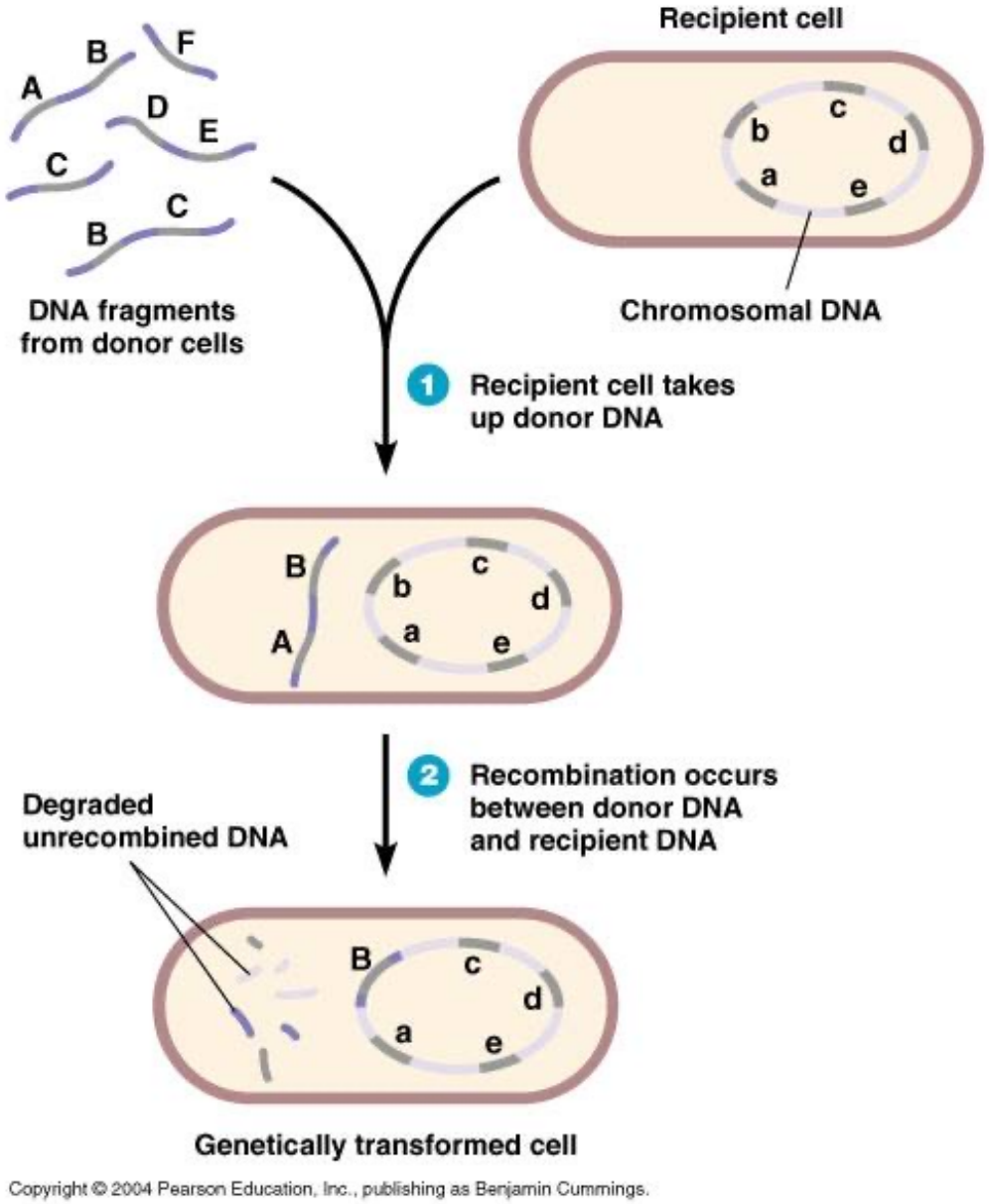
about-ecoli.com

Mycobacterium MLST

- Scheme devised for *Mycobacterium* genus
- Success among non-tuberculous Mycobacteria (NTMs)
- Limited genomic diversity in *Mycobacterium tuberculosis* means it is not a useful technique in this species (doi.org/10.1016/j.mimet.2013.01.020)
- Figure generated from all *Mycobacterium tuberculosis* data in PubMLST



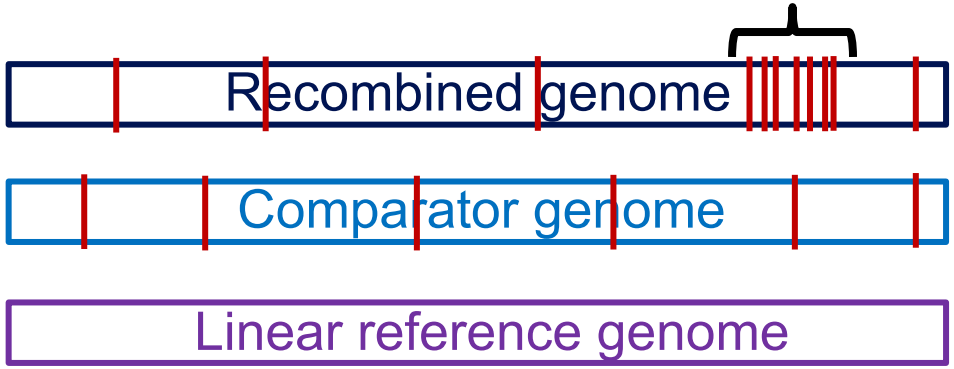
Recombination can confound typing



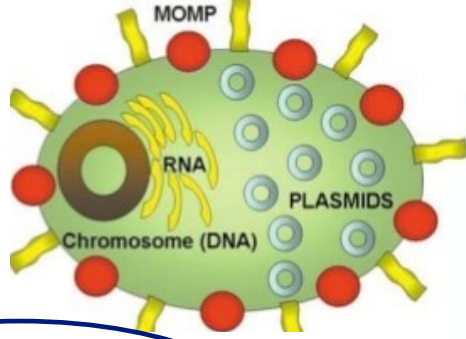
How to identify this is genome sequence:

Mutations (SNPs) often accumulate randomly

High density of SNPs implies gain of "foreign" DNA



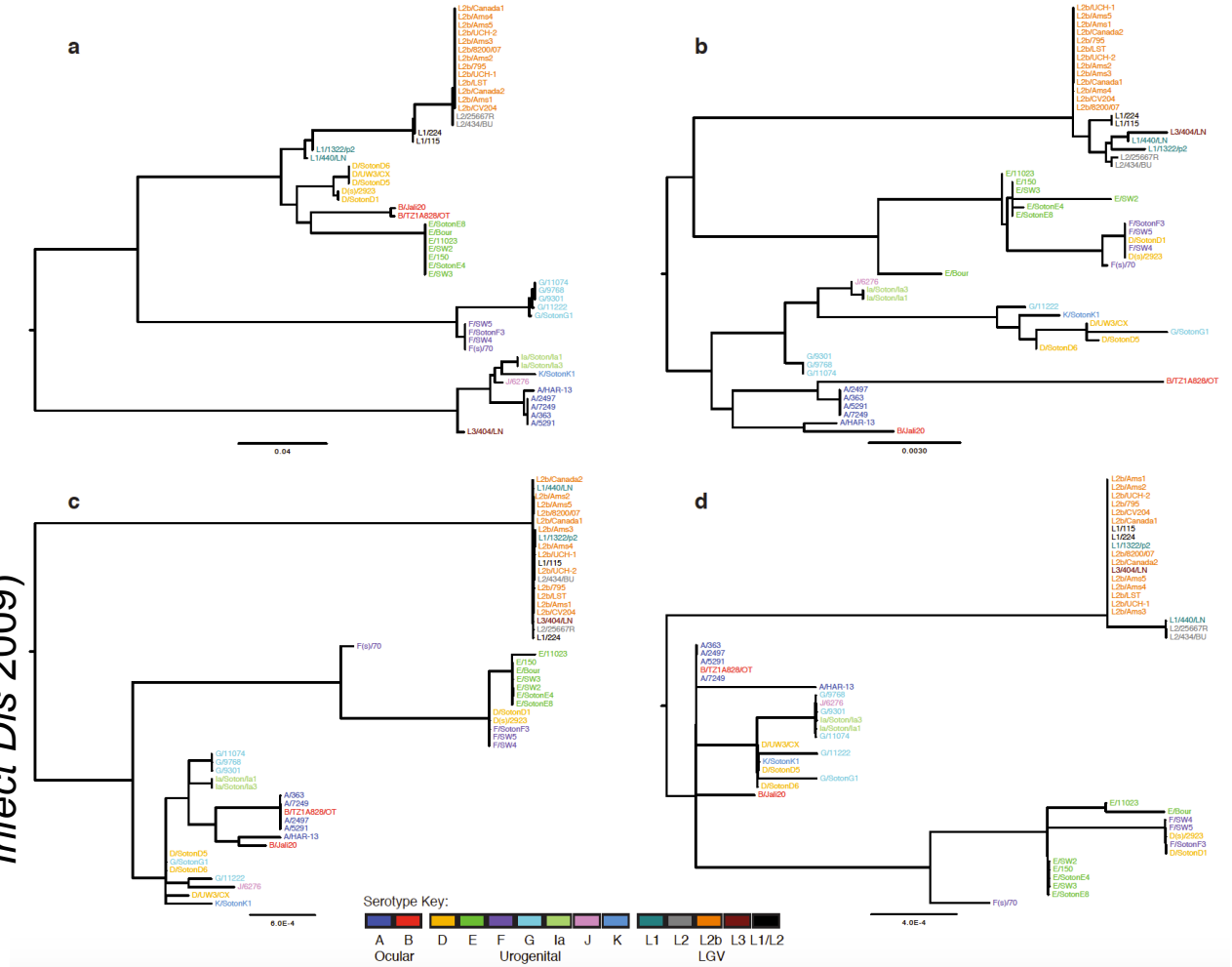
Recombination can confound typing: *Chlamydia trachomatis*



Colour mixing (tip name labels) shows how MOMP / *ompA* typing disagrees with full genome data
Tree structures show how resolution differs

ompA gene

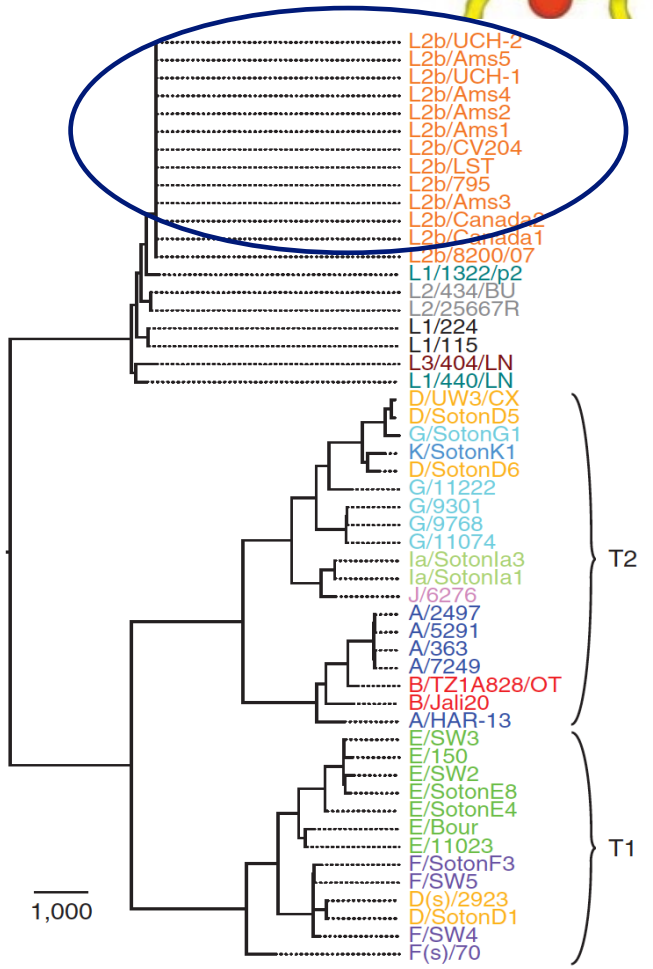
MLST (Dean et al. *Emerg Infect Dis* 2009)



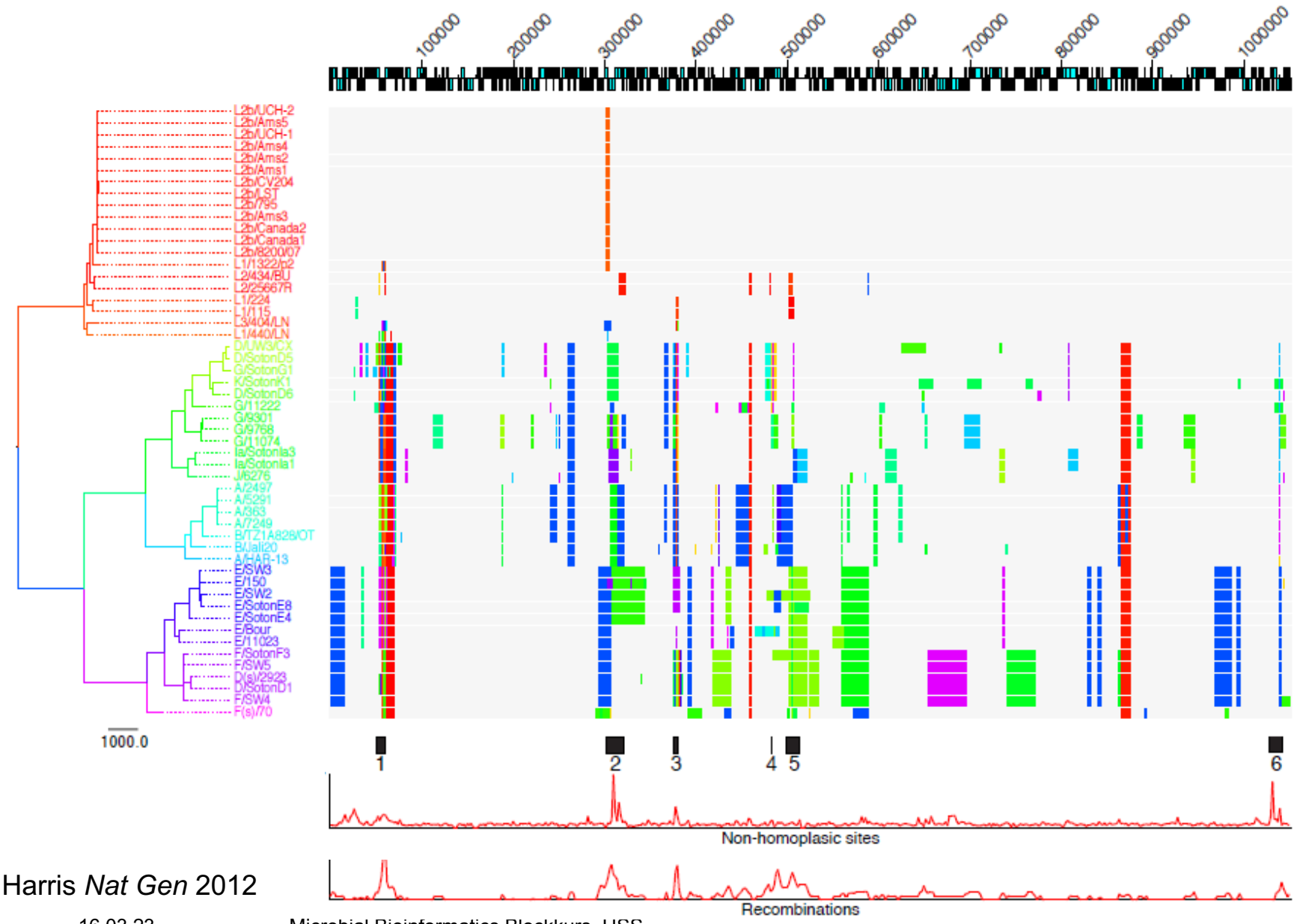
MLST (Klint et al. *J Clin Microbiol* 2007)

MLST (Pannekoek et al. *BMC Microbiol* 2008)

Genomic phylogeny



Recombination in *C. trachomatis*



Harris *Nat Gen* 2012

Large scale recombination:

- 26% of sites (4492 /17163) are homoplasic

Certain loci are more variable than others

Discussion

- Conventional typing was based on phenotypes
- Almost all typing now is based on DNA analysis
- The more information included, the more accurate
- MLST is based on a few conserved loci per species: can be analysed by PCR and Sanger sequencing or from the whole genome
- The nomenclatures behind typing are pervasive, especially among clinicians

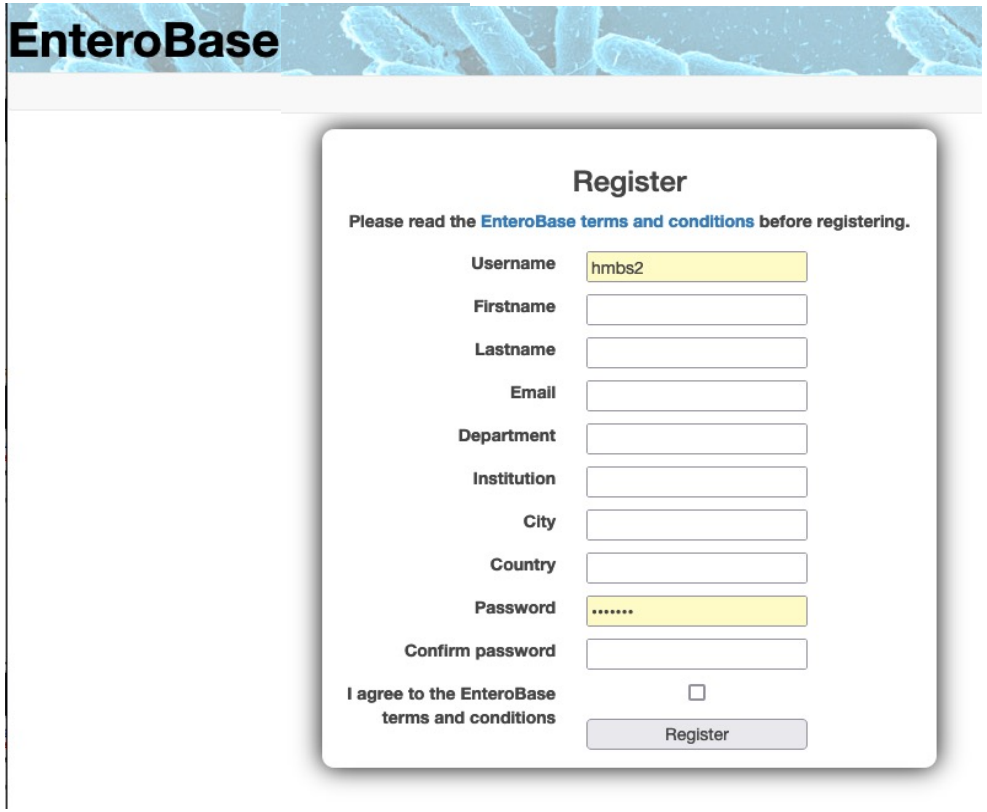
- Stay tuned for more information on genome-based typing tomorrow!

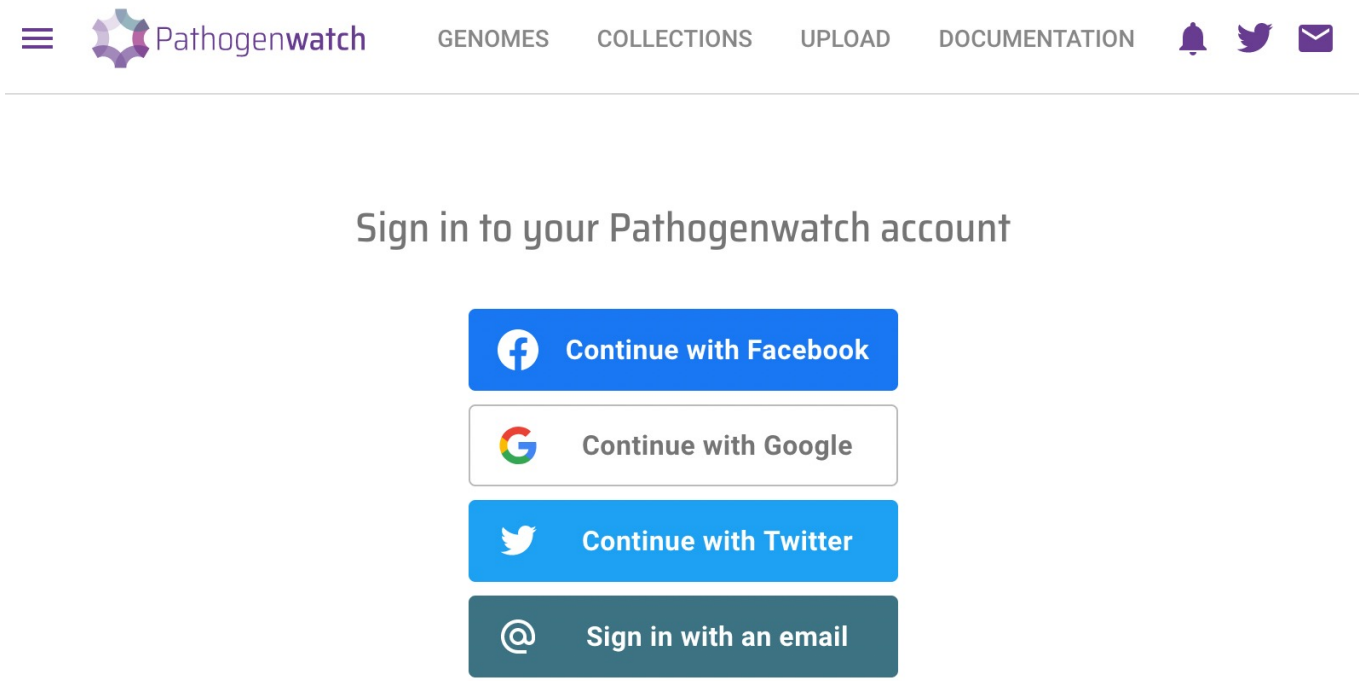
Small exercise

Please generate logins, for tomorrow's work, at:

Enterobase : https://enterobase.warwick.ac.uk/auth/register

PathogenWatch: https://pathogen.watch/sign-in





Reading suggestions

Introduction to microbial bioinformatics

https://www.sciencedirect.com/science/article/pii/S1198743X17307097?via%3Dihub

Recommended review

Balloux: From Theory to Practice: Translating Whole-Genome Sequencing (WGS) into the Clinic, 10.1016/j.tim.2018.08.004

rMLST:

https://doi.org/10.1099/mic.0.055459-0

PubMLST and BIGSdb:

https://wellcomeopenresearch.org/articles/3-124/v1

Enterobase

https://genome.cshlp.org/content/30/1/138.long

Many thanks for your attention

Questions??