

# Know your enemy: a primer on the population biology and molecular epidemiology of bacterial pathogens

Martin Maiden  
Department of Biology



UNIVERSITY OF  
**OXFORD**

# Learning Outcomes

To appreciate:

1. The relevance of molecular approaches to questions in clinical microbiology.
2. The nature and extent of bacterial diversity both within and among genomes.
3. How sequence data illuminate bacterial diversity.
4. The variety of bacterial population structures and the implications of horizontal gene transfer (HGT).
5. Practical implications of molecular epidemiology and population biology.

**Genomics and Diagnostics Meeting**  
**31st March – 2nd April 2003**  
**Hinxton Hall Conference Centre, Hinxton, UK**

**Monday 31st March**

17:00	Arrival and registration	<i>Conference Centre</i>
18:30	Introduction and Welcome <i>Dr Karen Kennedy</i>	<i>James Watson Pavillion</i>
A global overview of the disease burden where diagnostics can make a difference <i>Sir David Weatherall, University of Oxford</i>		
19:30	Pre-dinner drinks	<i>Hall Foyer</i>
20:00	Dinner	<i>Restaurant</i>

**Tuesday 1st April**

<b>0830 – 1030</b>	<b>Session I: ‘Setting the scene: Where are we today?’</b> <b>Chair - Professor Martin Maiden</b>	<i>James Watson Pavillion</i>
--------------------	--	-------------------------------

ADVANCED COURSES  
GENOME CAMPUS

wellcome  
trust

Genomics and  
Clinical  
Microbiology

6-11 February 2005

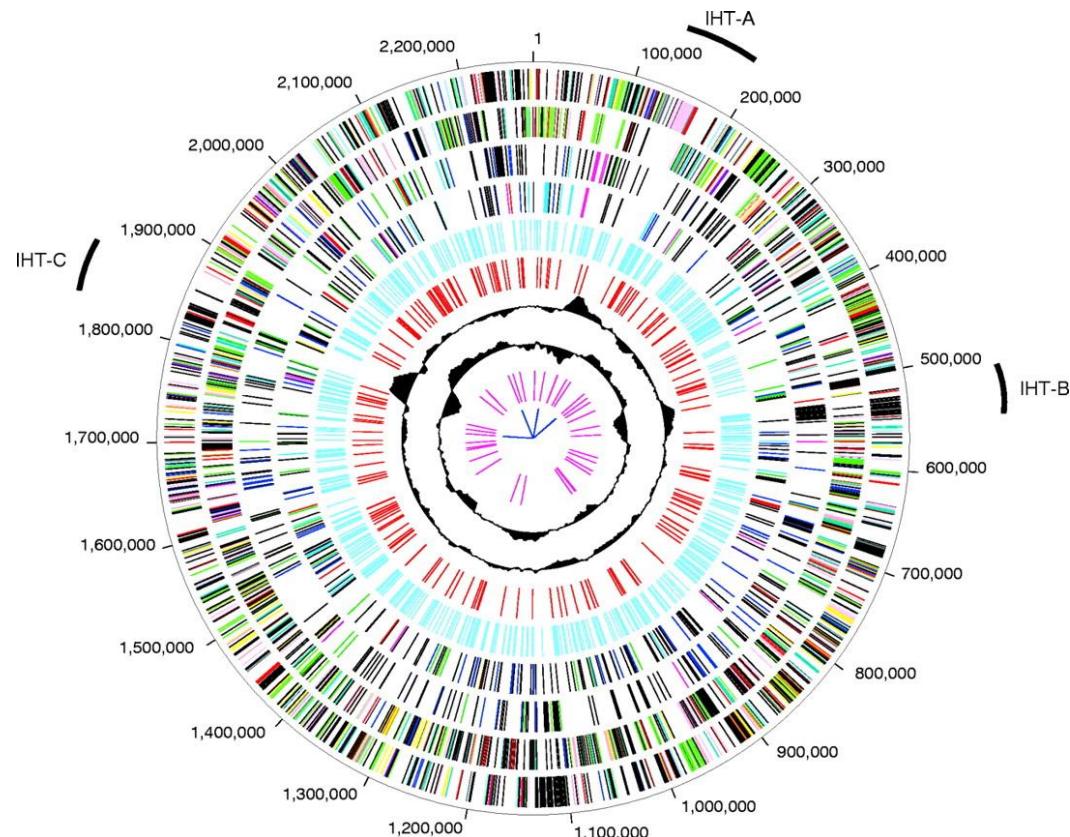
Genetic techniques for  
medical microbiology



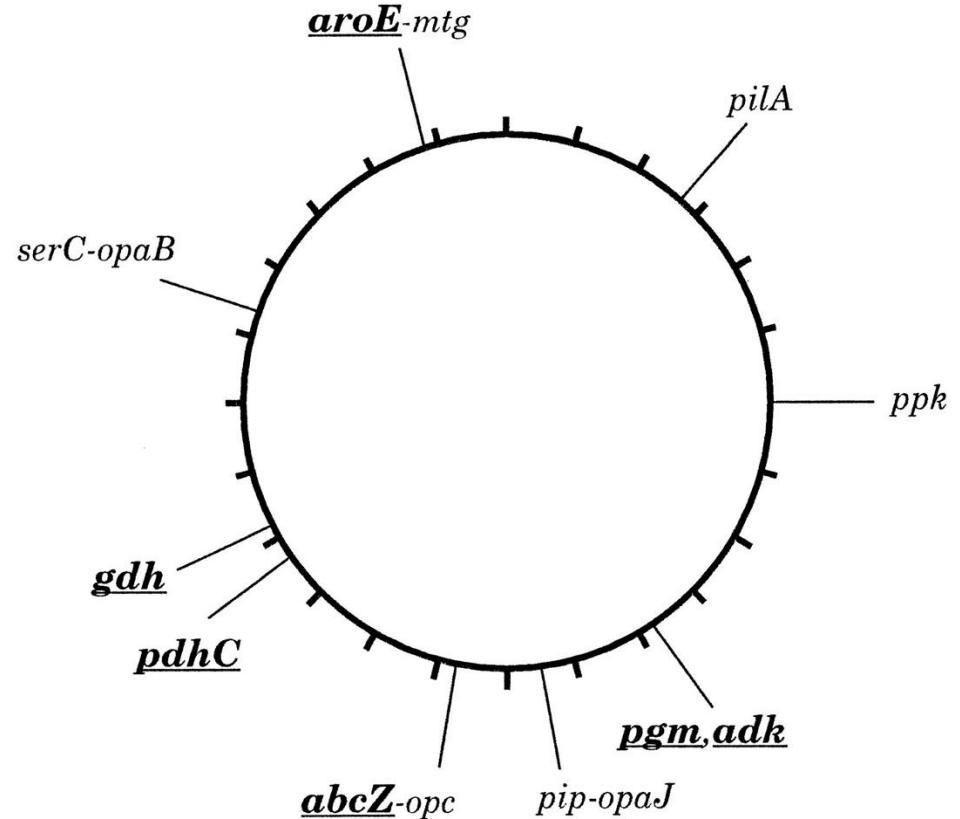


# A quarter of a century of:

## Genomes



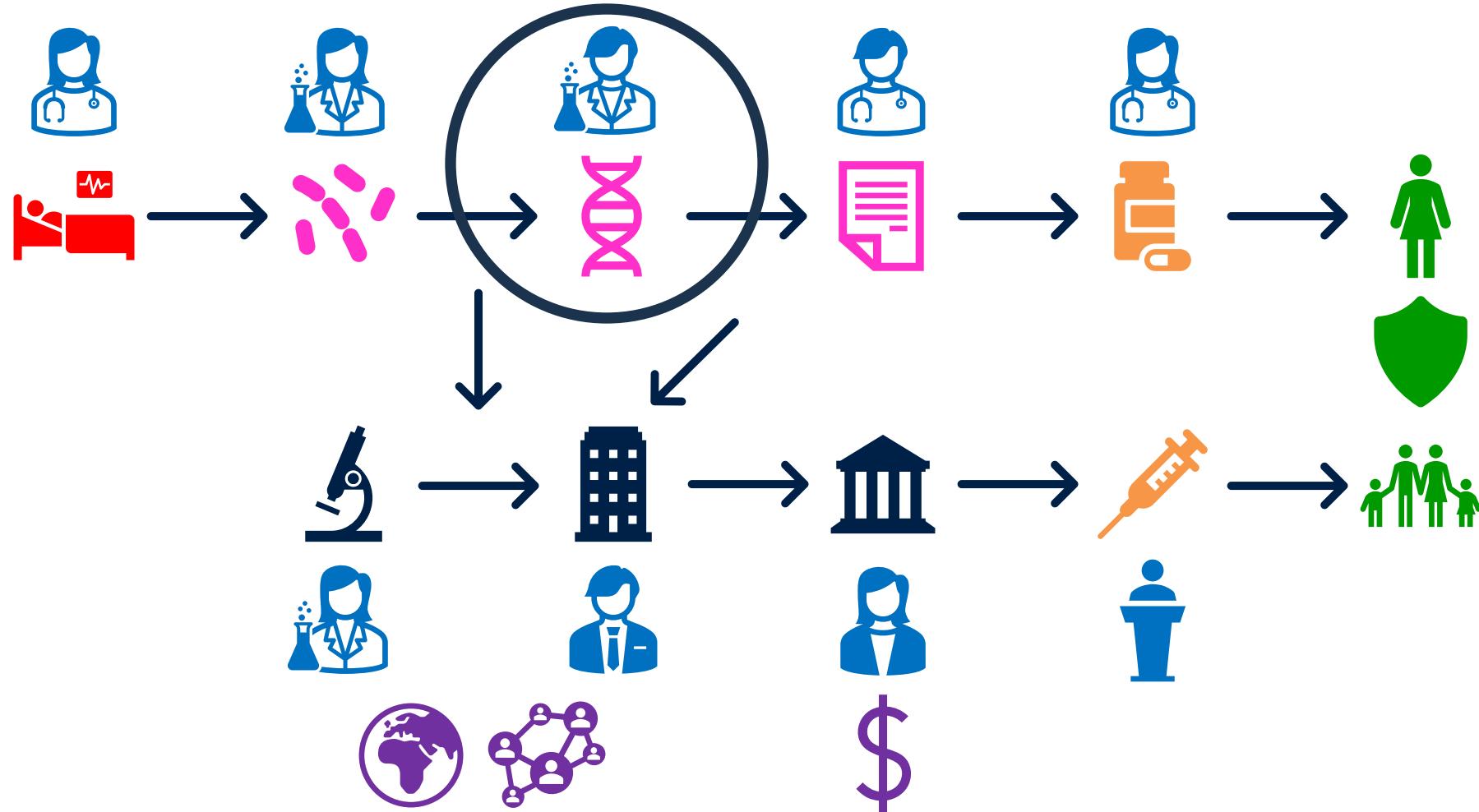
## Population genetics



Tettelin, H., Saunders, N. J., et al., Moxon, E. R., Rappuoli, R. & Venter, J. C. (2000). Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**, 1809-1815.

Maiden, M. C. J., Bygraves, J. A., et al., Feavers, I. M., Achtman, M. & Spratt, B. G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *P Natl Acad Sci USA*. **95**, 3140-3145.

# Genomics in Clinical and Public Health Microbiology



# Questions in clinical microbiology

## evolution

# Centuries+

# emergence

decades

years

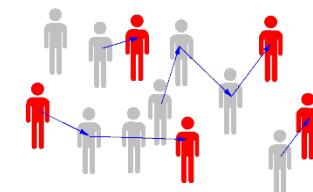
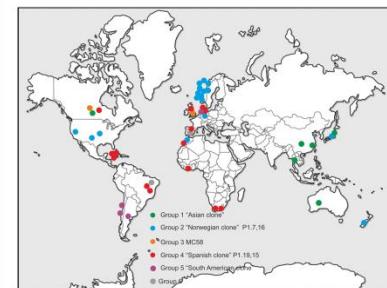
# epidemiology

## months

## weeks

# diagnosis

hours



# High

**Low**

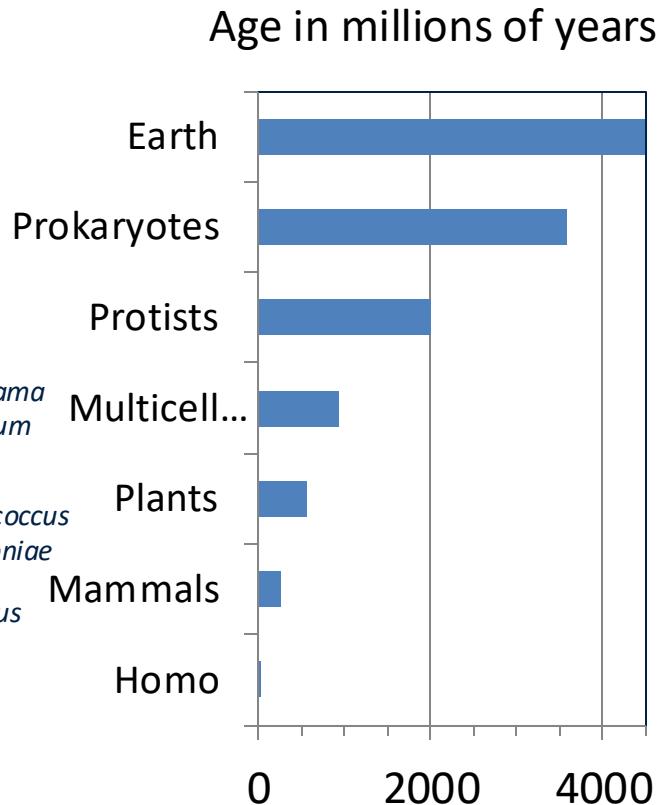
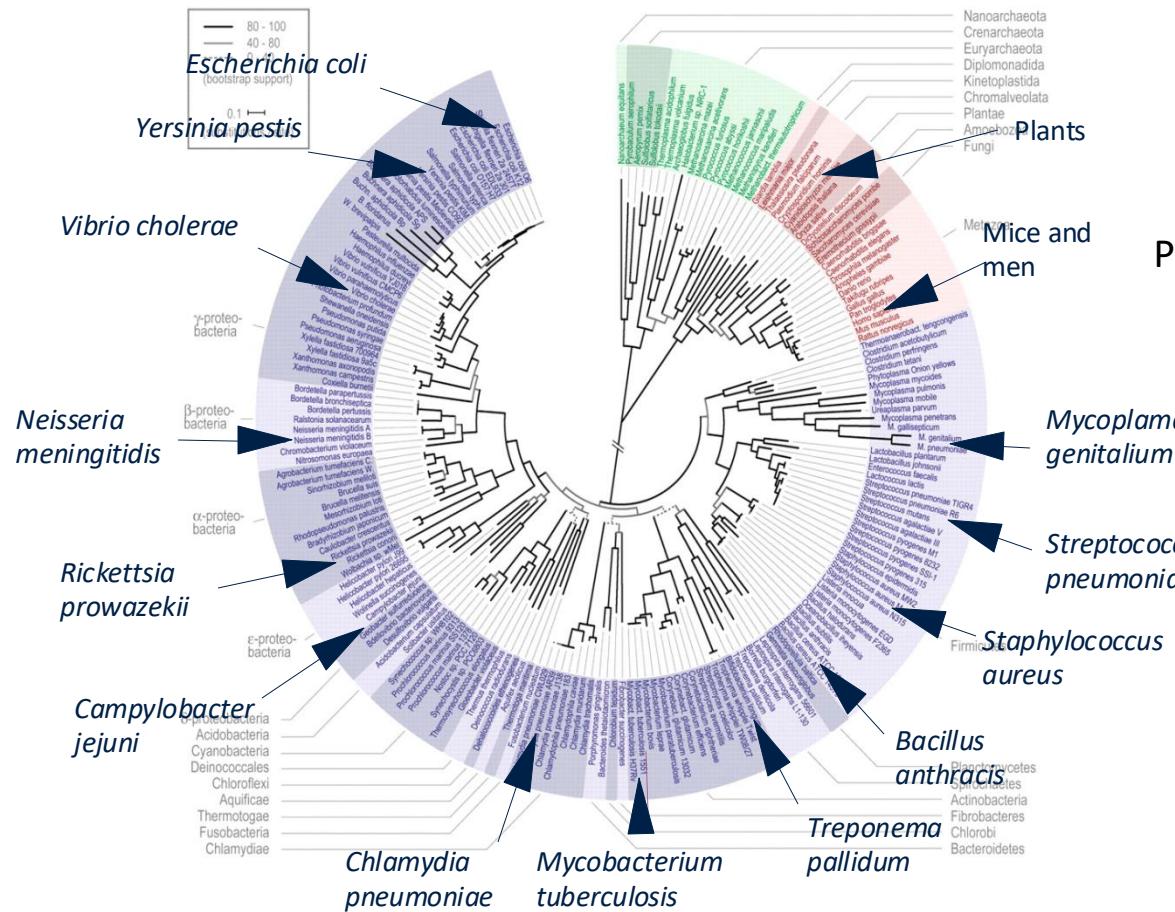
## Relative amount of genetic change

**Low**

High

## Relative discrimination required

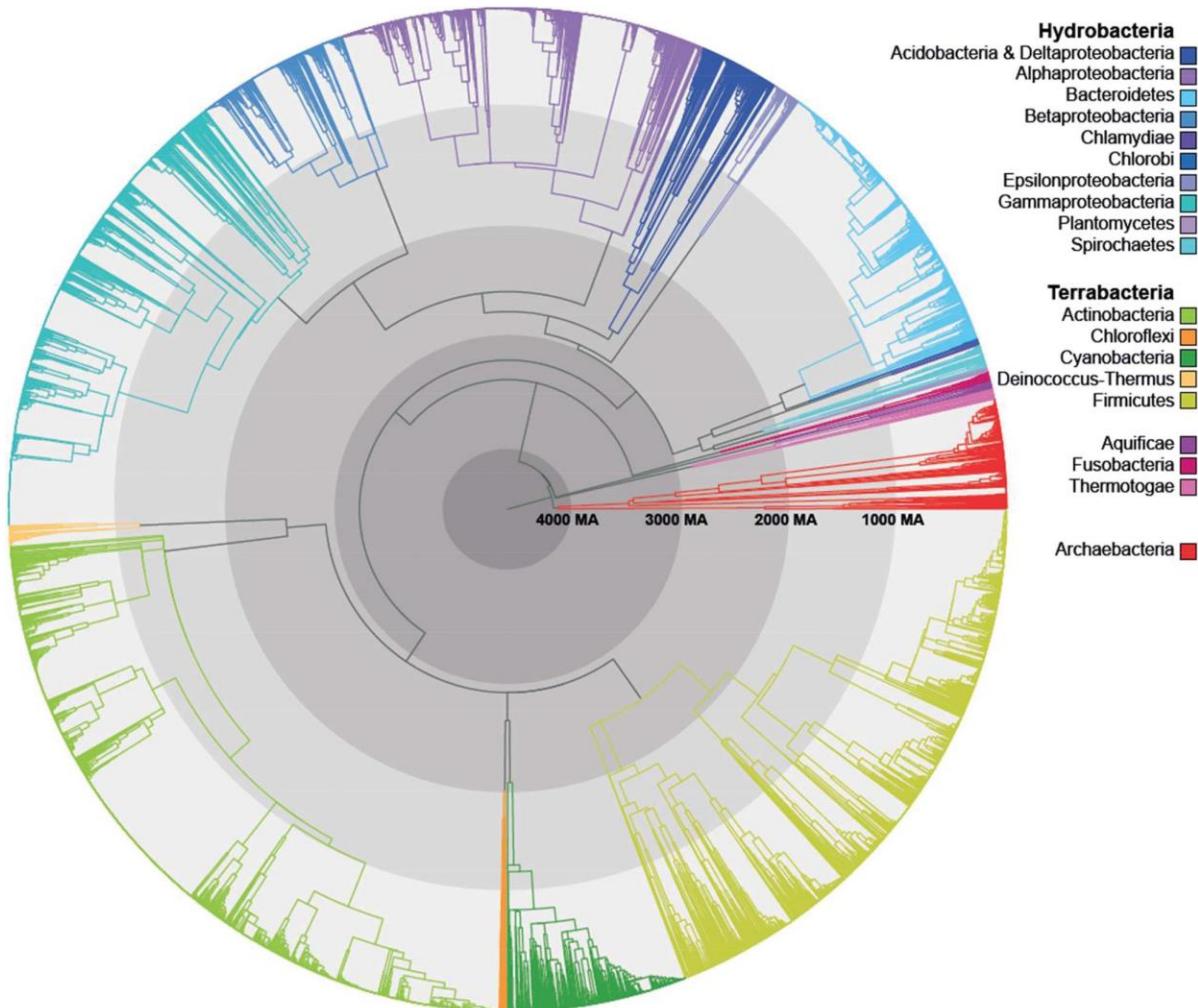
# The scale of the problem: planet of the bacteria



Stephen Jay Gould, "Planet of the Bacteria," Washington Post Horizon, 1996, 119 (344)

Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283-1287.

# Prokaryote time-tree for 11,784 ‘species’ estimated from small subunit ribosomal RNA sequences



Marin, J., Battistuzzi, F. U., Brown, A. C. & Hedges, S. B. (2017). The Timetree of Prokaryotes: New Insights into Their Evolution and Speciation. *Mol Biol Evol.* **34**, 437-446.

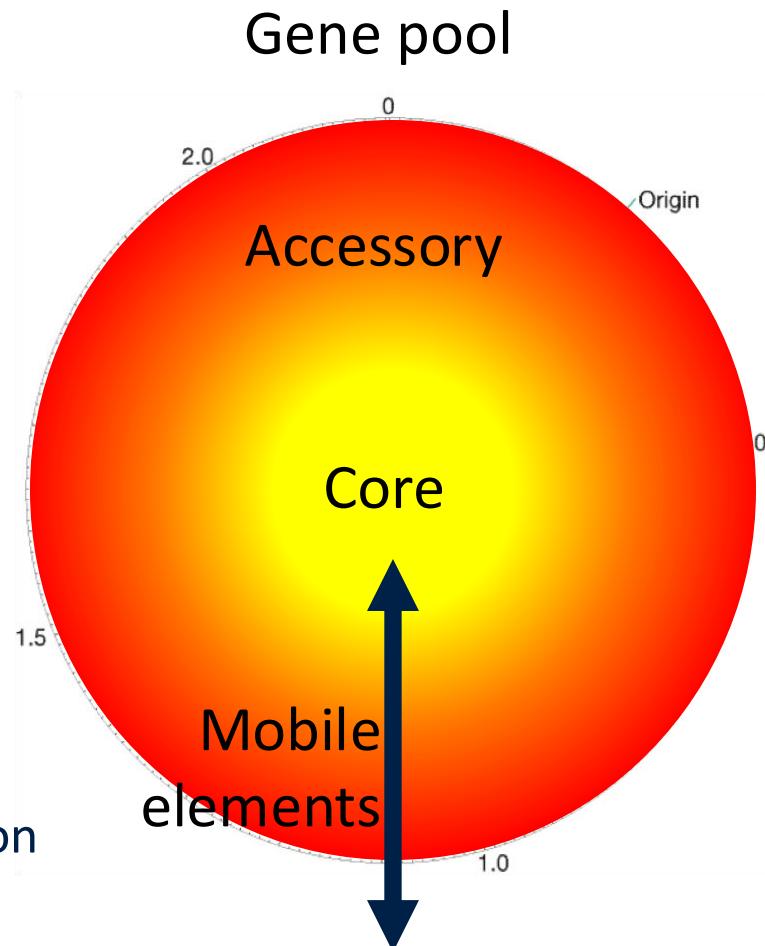
# Bacterial genome diversity

## Core genome:

e.g. DNA replication, ribosomes, cell envelope, key metabolic pathways.

## Parasitic elements (phages plasmids)

e.g. toxins, restriction/modification systems.



## Accessory genome:

e.g. alternative metabolic pathways, transport systems.

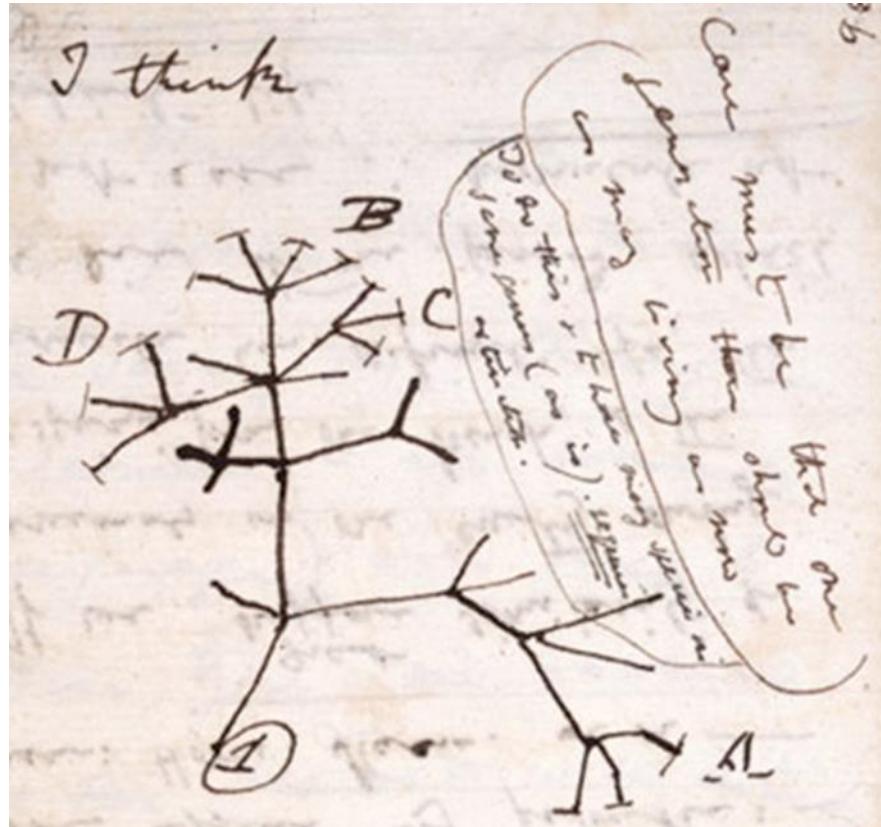
Tettelin, H., Riley, D., Cattuto, C. & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology* 11, 472-477.

## Gene pool:

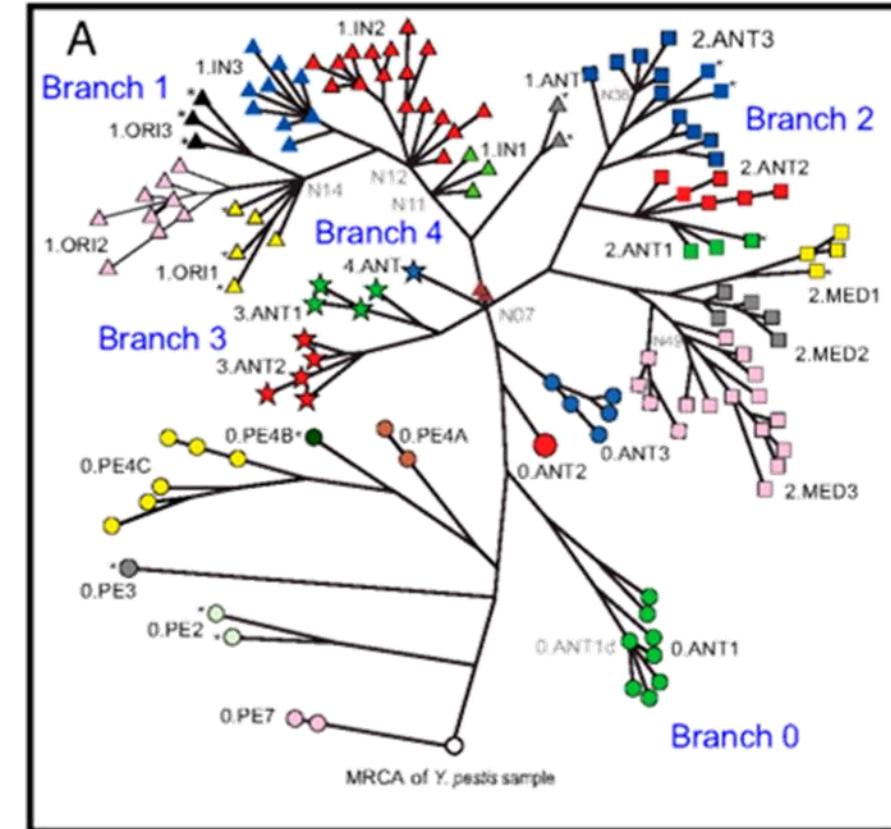
e.g. antibiotic resistance, degradative metabolism.

Genetic elements may be subject to **stabilising (negative)** or **diversifying (positive)** selection or be **neutral** (rare in most bacteria).

# I think ... of trees?



Darwin, C.R., Notebook B



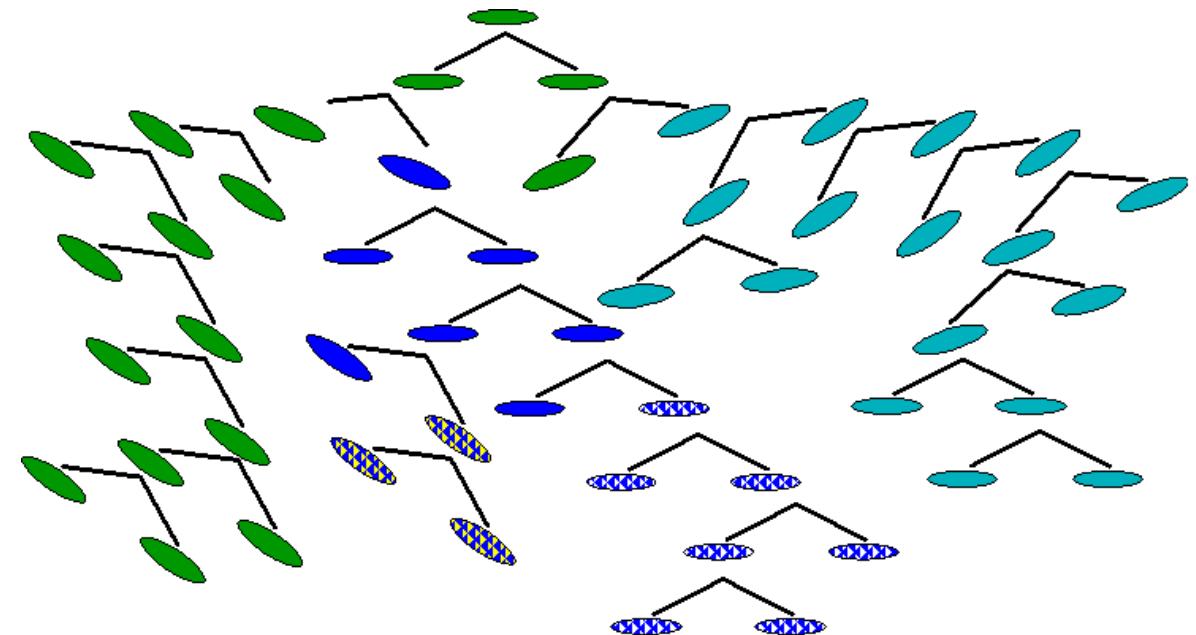
Cui, Y., et al. (2013). Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natil. Acad Sci USA* **110**, 577-582.

# Patterns in sequence variation

## a primer on bacterial population biology

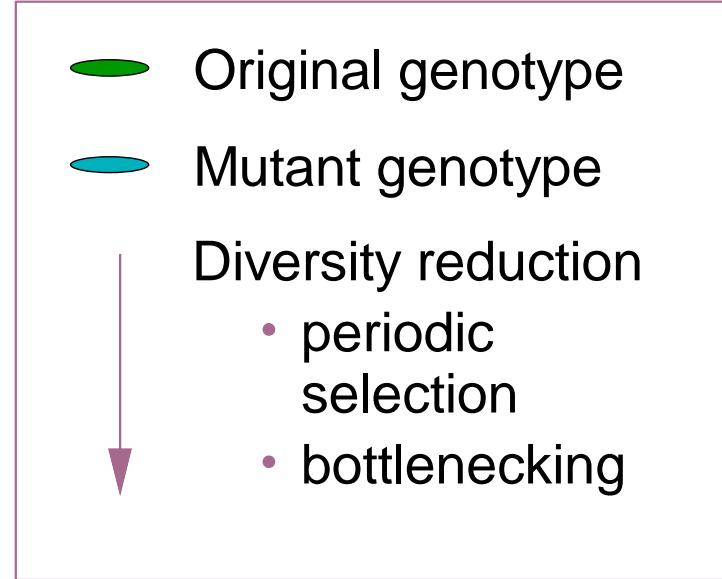
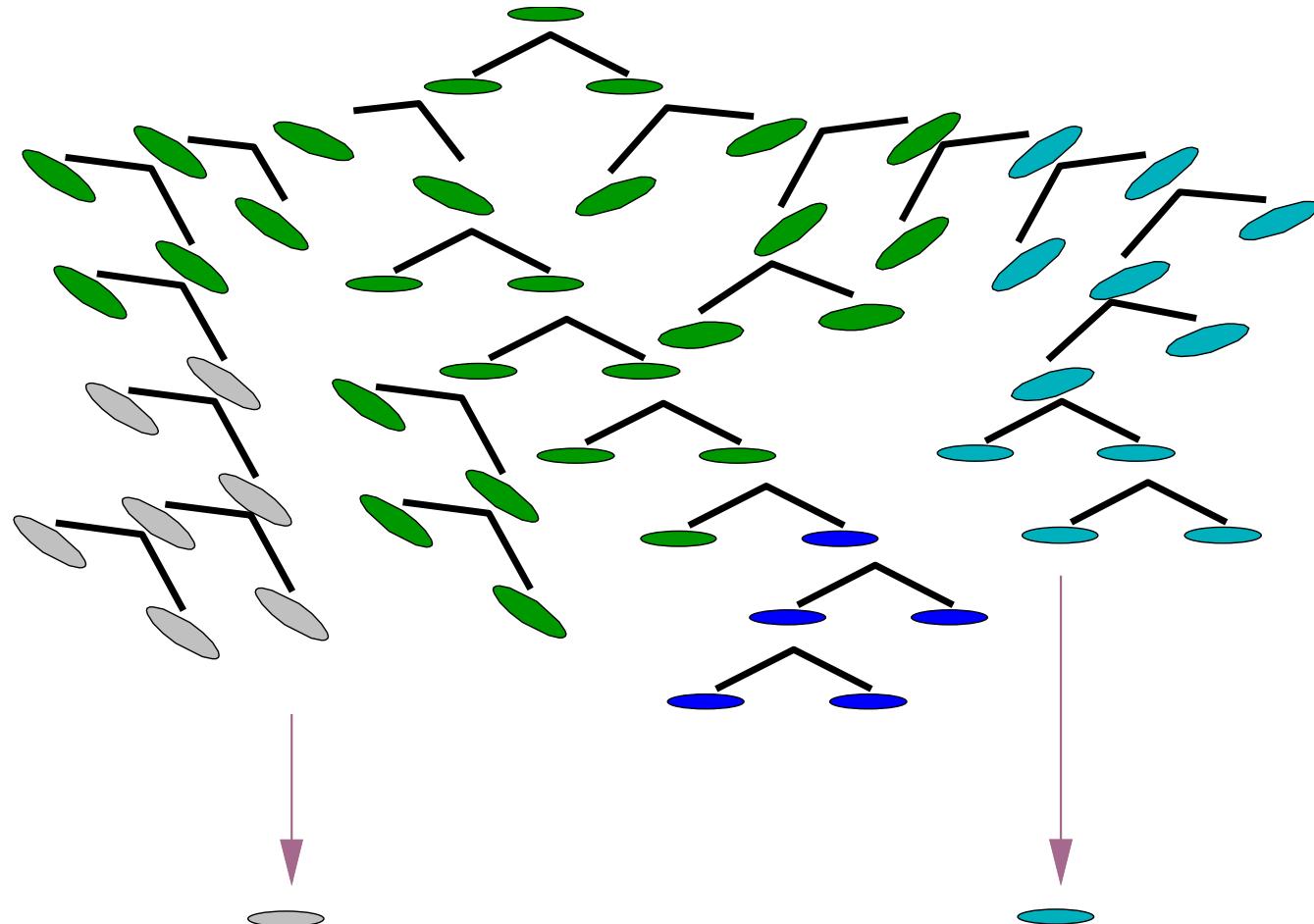
Ideas about bacterial populations are dominated by the facts that bacteria:

- are asexual;
- reproduce by binary fission, with each ‘mother’ cell giving rise to two identical ‘daughter’ cells (**clones**);
- accumulate genetic change by ‘vertical’ inheritance.



Gupta, S. & Maiden, M.C.J. (2001). Exploring the evolution of diversity in pathogen populations. *Trends in Microbiology* 9, 181-192.

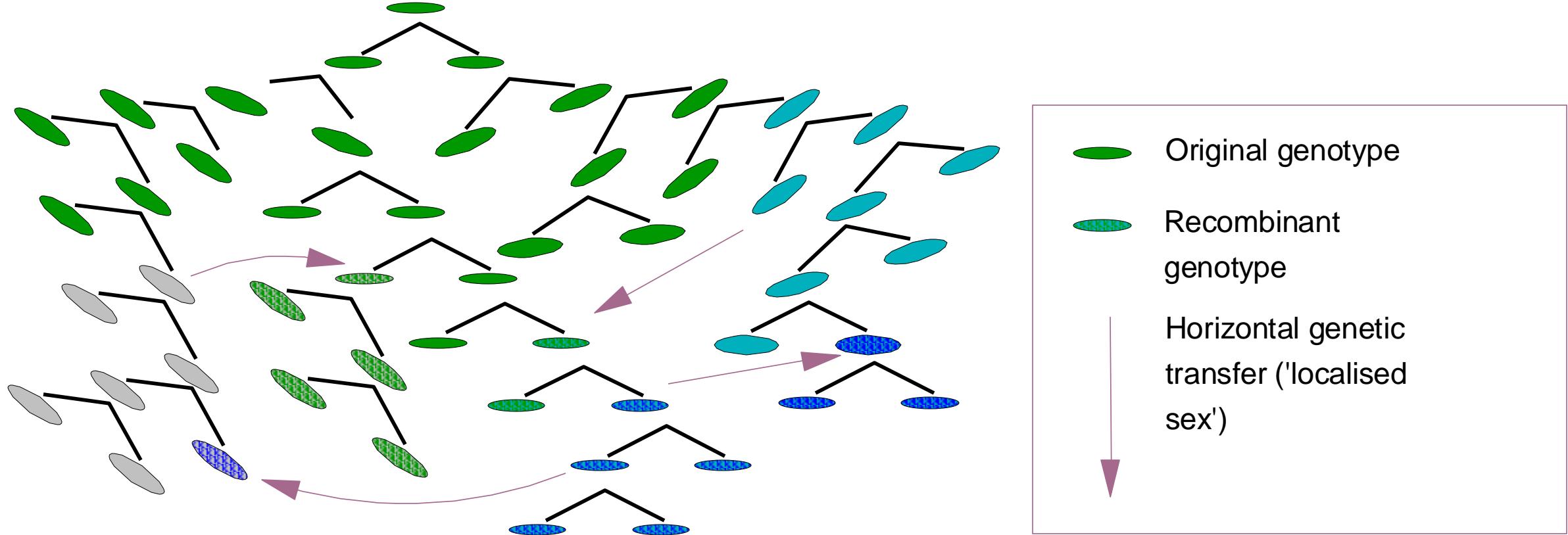
# The clonal population model: asexuality with diversity reduction



**Levin BR.** 1981. Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. *Genetics* 99(1):1-23.

Within this model, Bacterial populations should be easy to understand ...

# Impact of Horizontal genetic exchange (HGT) on bacterial population structure



... however, HGT disrupts clonal structure, breaking down tree-like phylogeny, linkage disequilibrium and congruence.

Maynard Smith J, Dowson CG, Spratt BG. (1991). Localized sex in bacteria. *Nature* **349**, 29-31.

# Clonal and non-clonal population structures

## Clonal

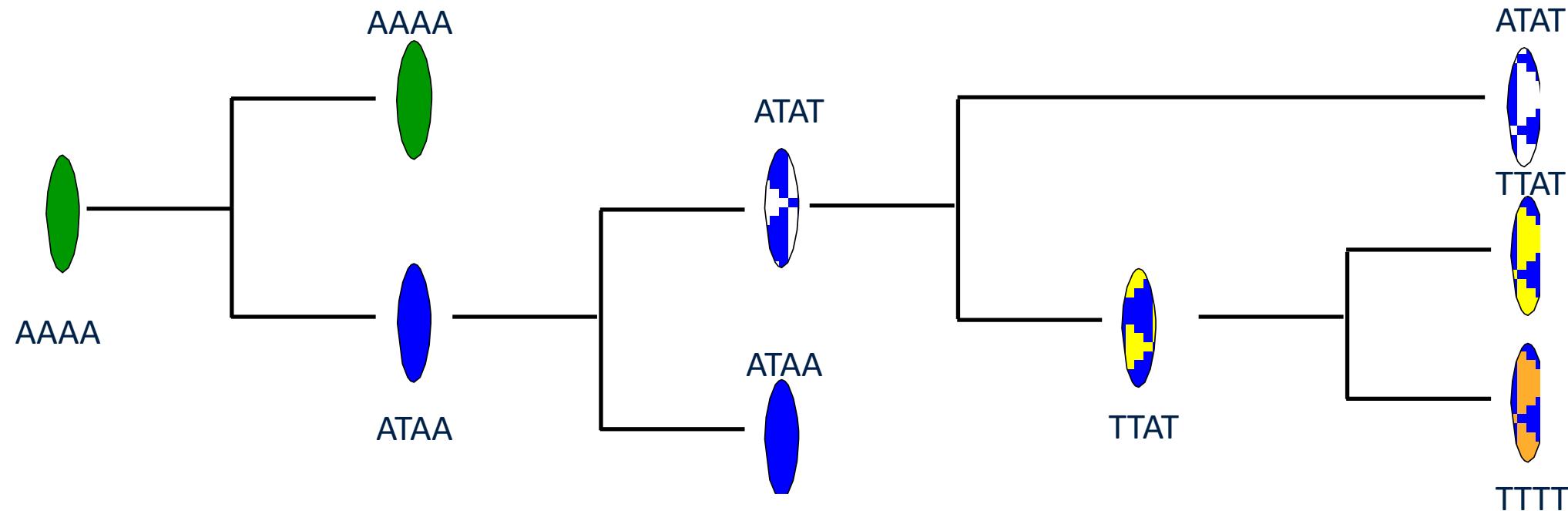
- Linkage disequilibrium
  - non-random allele combinations.
- Tree-like phylogeny
  - a bifurcating tree accurately models descent.
- Congruence
  - the same phylogenetic signal is recorded throughout the genome.

## Non-clonal

- Linkage equilibrium
  - random allele combinations.
- Net-like phylogeny
  - a bifurcating tree cannot model descent.
- Incongruence
  - different phylogenetic signals are recorded throughout the genome.

# Molecular epidemiology made easy: clonality

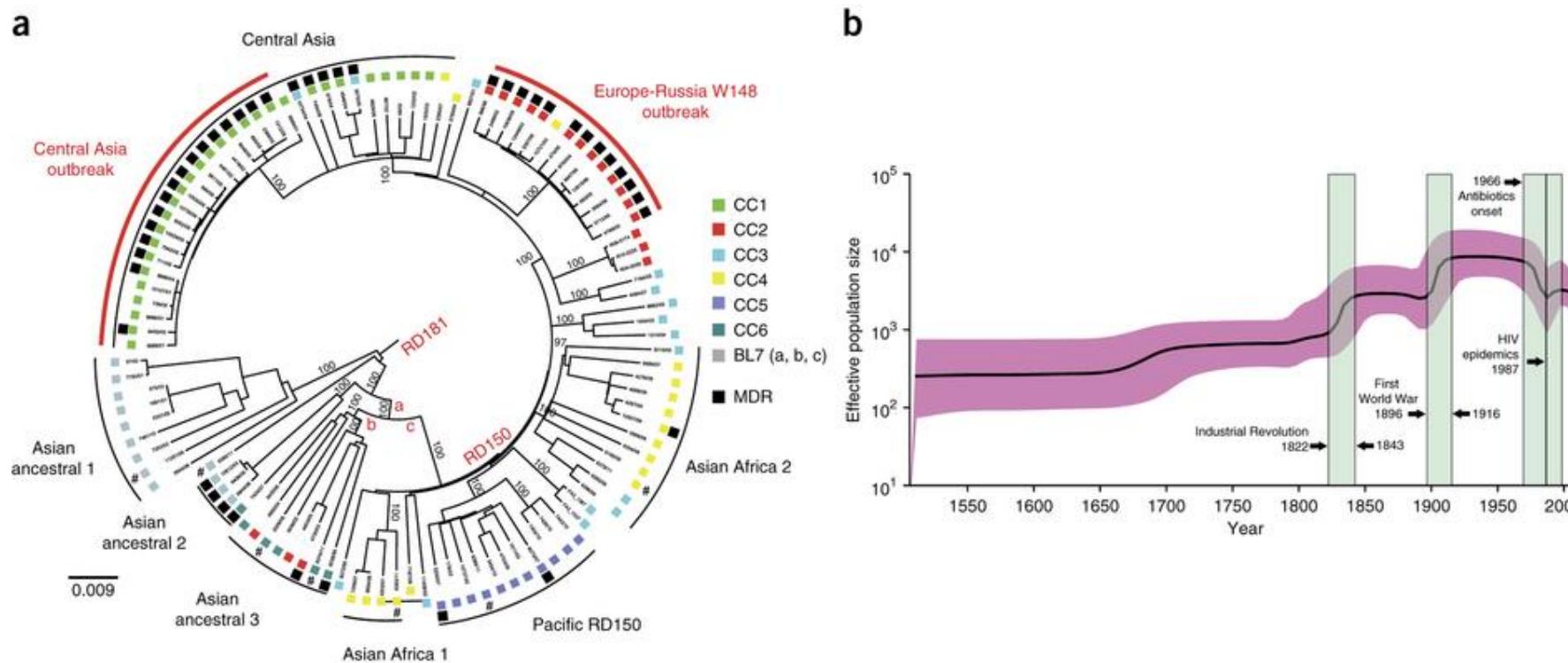
Centuries+      decades      years      months      weeks      days      hours



Progressive accumulation of genetic change

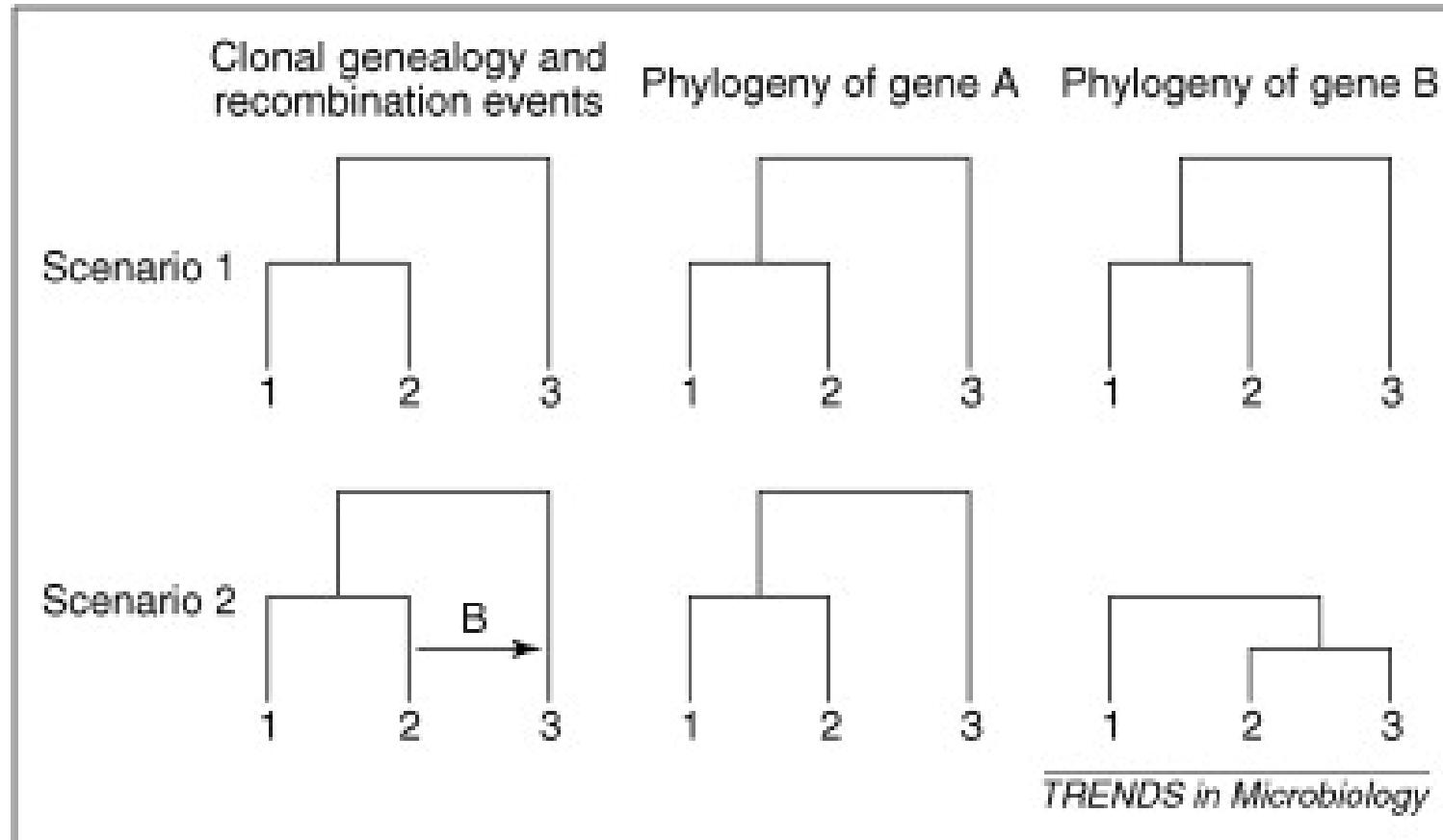


# *Mycobacterium tuberculosis* is clonal: Spread of the Beijing lineage



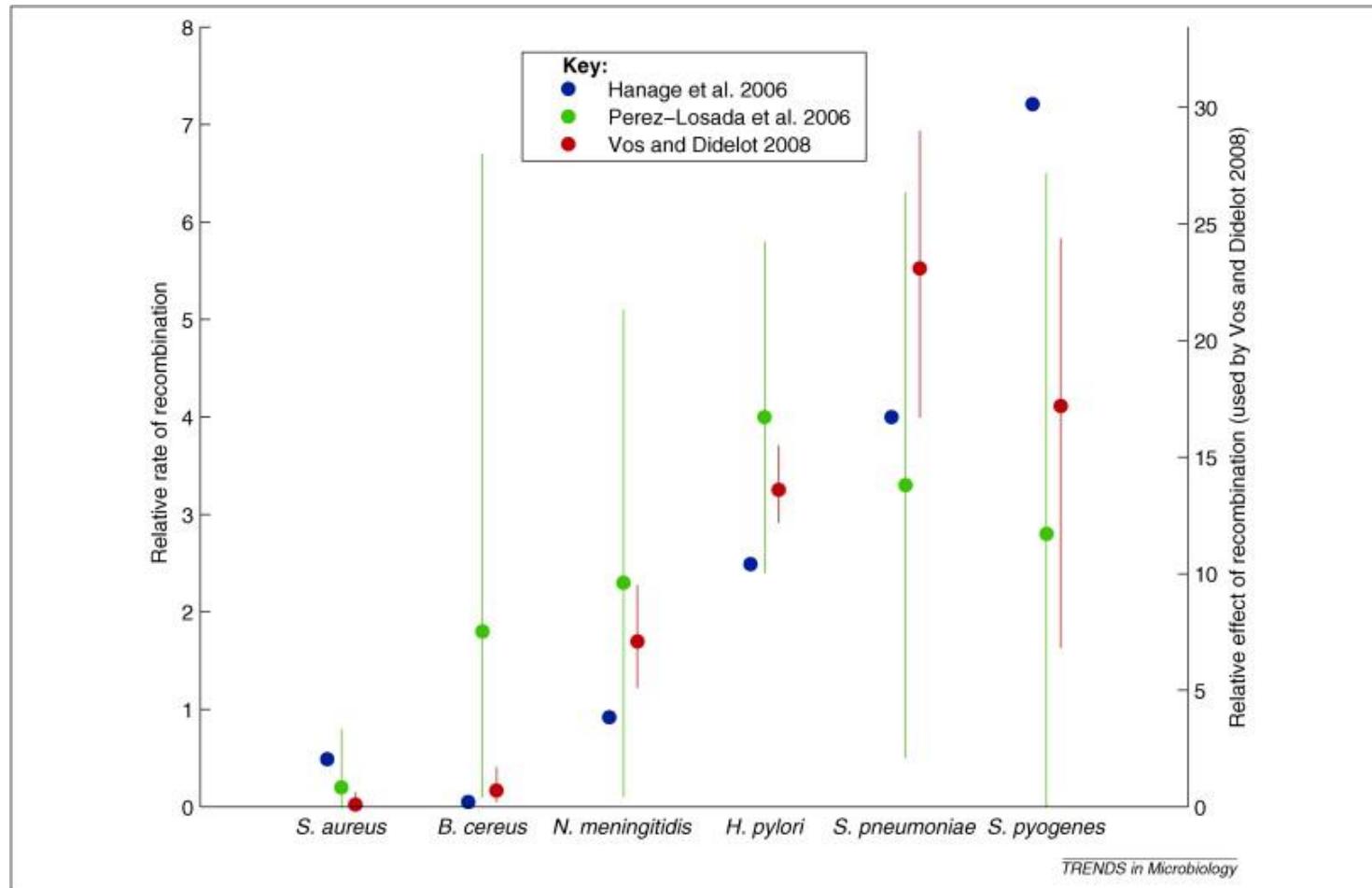
**Merker, M., Blinet. Al. (2015).** Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* **47**, 242-249.

# Recombination and phylogeny



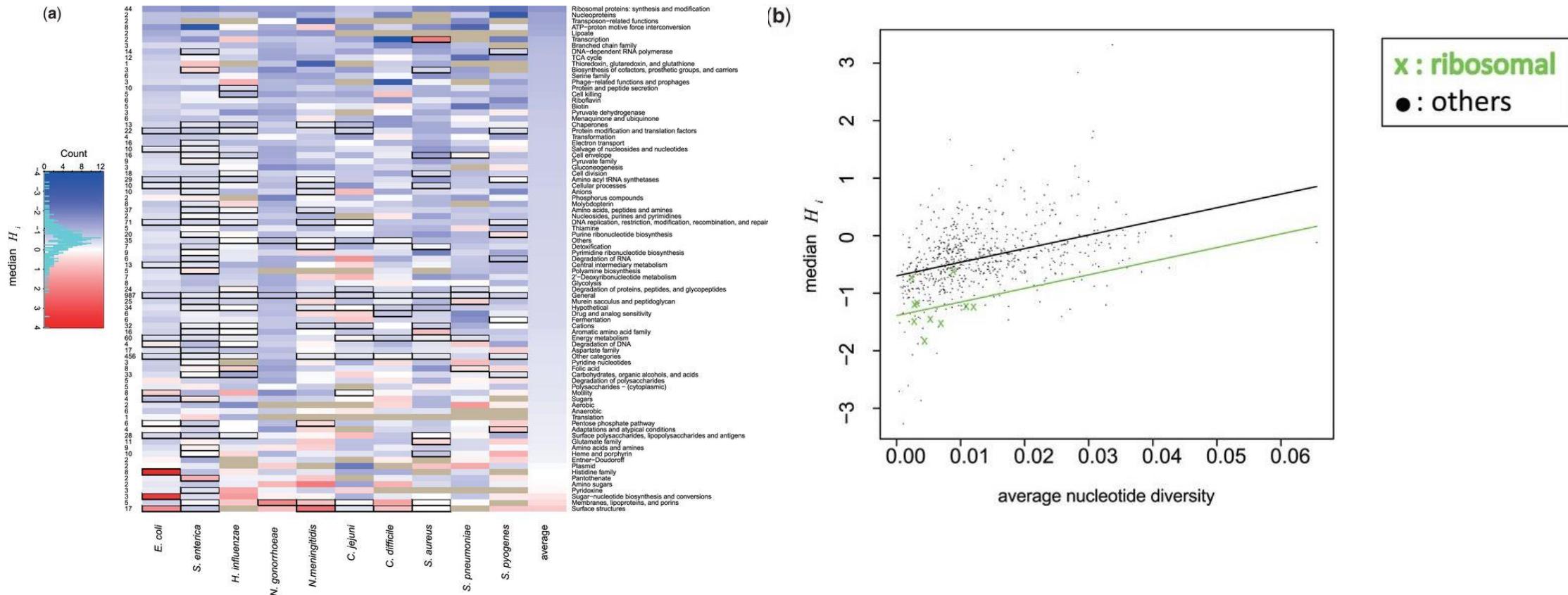
Didelot, X. & Maiden, M. C. (2010). Impact of recombination on bacterial evolution. *Trends in Microbiology* 18, 315-322.

# Rates of recombination are not uniform among bacteria



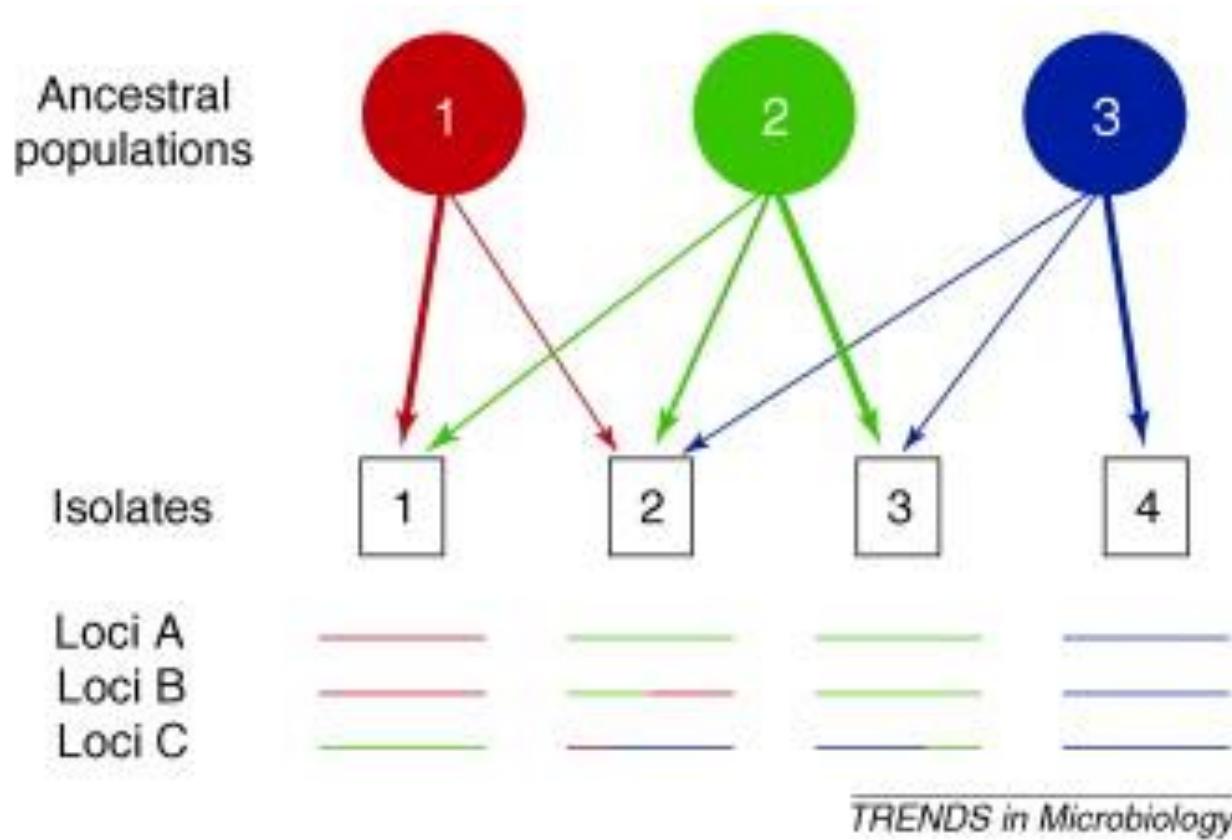
**Didelot, X. & Maiden, M. C. (2010).** Impact of recombination on bacterial evolution. *Trends in Microbiology* 18, 315-322.

# Rates of recombination are not uniform within the genome



**Yahara, K., Didelot, X., Jolley, K. A., Kobayashi, I., Maiden, M. C., Sheppard, S. K. & Falush, D. (2016). The Landscape of Realized Homologous Recombination in Pathogenic Bacteria. *Mol Biol Evol.* **33**, 456-471.**

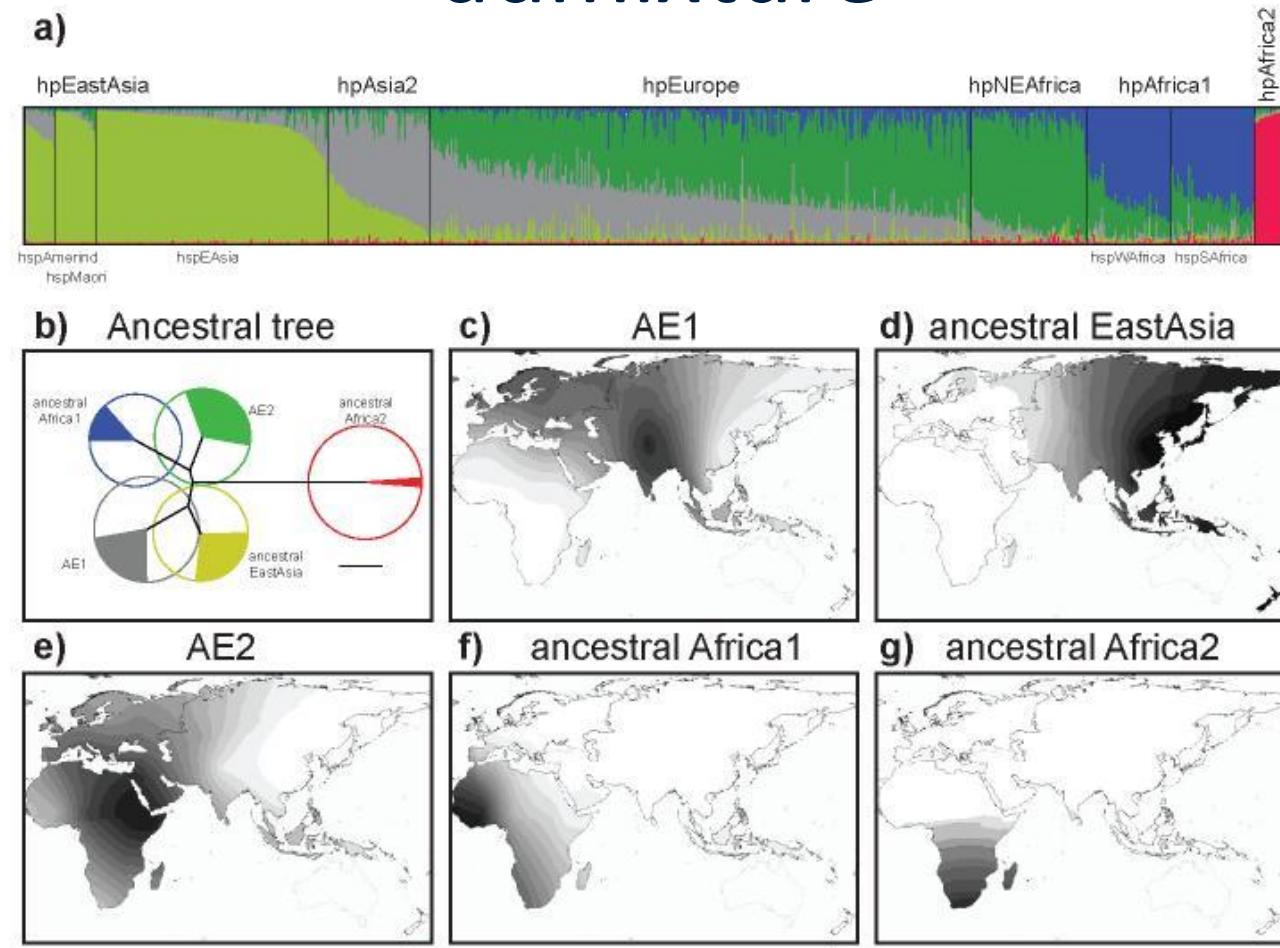
# Population admixture: the STRUCTURE model



The top part shows the ancestral populations in distinct colors. The bottom part shows the genotypes of the isolates, with fragments colored according to their ancestral population of origin.

**Didelot, X. & Maiden, M. C. (2010). Impact of recombination on bacterial evolution. *Trends in Microbiology* 18, 315-322.**

# *Helicobacter pylori* is non-clonal: phylogeography & admixture



Linz, B., F. Balloux, Y. Moodley, A. Manica, H. Liu, P. Roumagnac, D. Falush, C. Stamer, F. Prugnolle, S. W. van der Merwe, Y. Yamaoka, D. Y. Graham, E. Perez-Trallero, T. Wadstrom, S. Suerbaum, and M. Achtman. (2007). An African origin for the intimate association between humans and *Helicobacter pylori*, *Nature*, 445: 915-8.

# Dealing with partial HGT

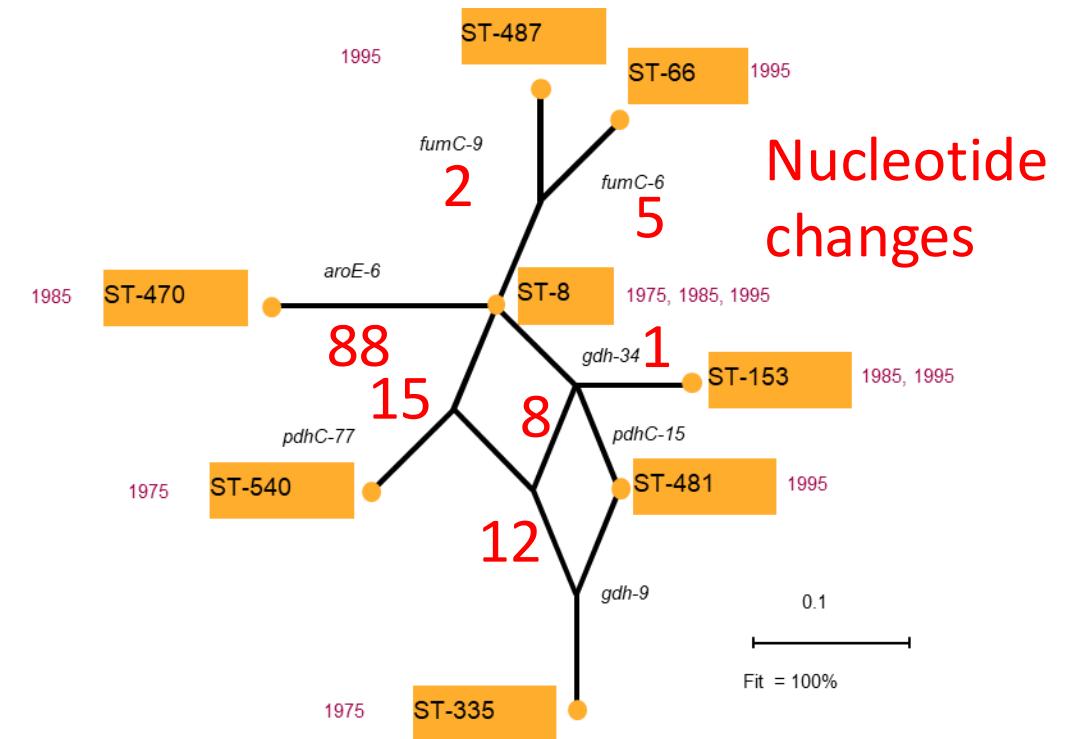
Recombination resulting from HGT violates the assumptions of clonal evolution. It can be accounted for in models of bacterial evolution by:

- Ignoring it, which only works if HGT is uncommon;
- Using loci and alleles as the unit of analysis (*e.g.* gene-by-gene MLST approaches);
- Identifying likely HGT and removing it (*e.g.* GUBBINS);
- Estimating relative contributions of recombination and mutation in evolutionary models (*e.g.* ClonalFrame ML).

# MLST: allele-based analyses, sequence types (STs) and clonal complexes (ccs)

ST	<i>adk</i>	<i>abcZ</i>	<i>aroE</i>	<i>fumC</i>	<i>gdh</i>	<i>pdhc</i>	<i>pgm</i>
8	2	3	7	2	8	5	2
66	2	3	7	6	8	5	2
153	2	3	7	2	34	5	2
335	2	3	7	2	9	15	2
470	2	3	6	2	8	5	2
481	2	3	7	2	9	5	2
487	2	3	7	9	8	5	2
540	2	3	7	2	8	77	2

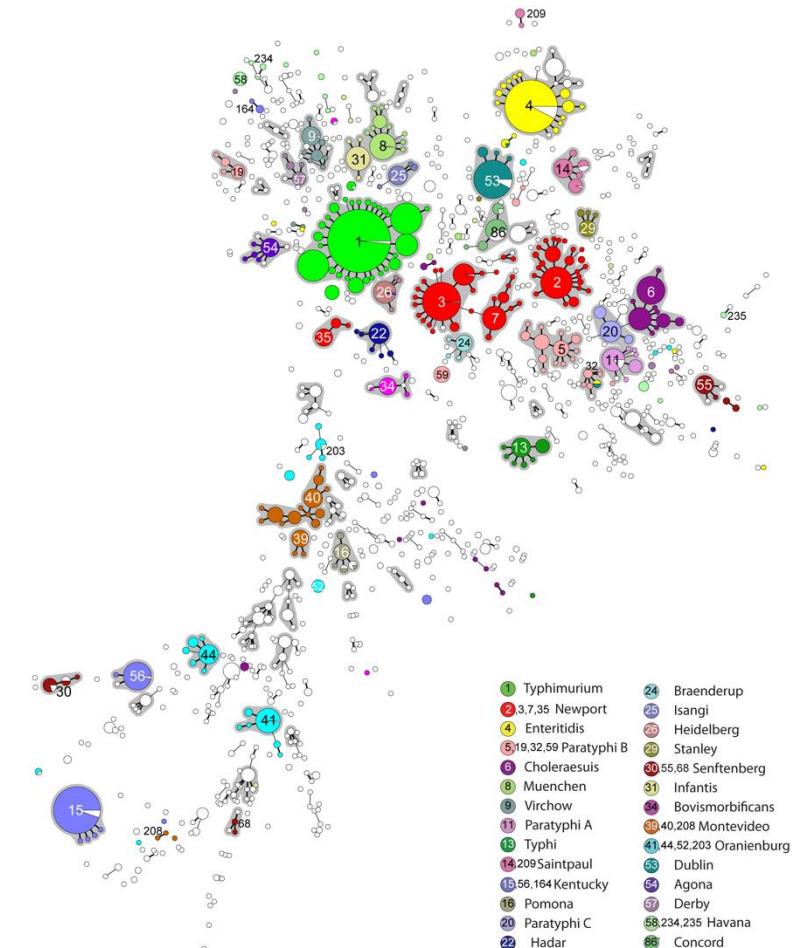
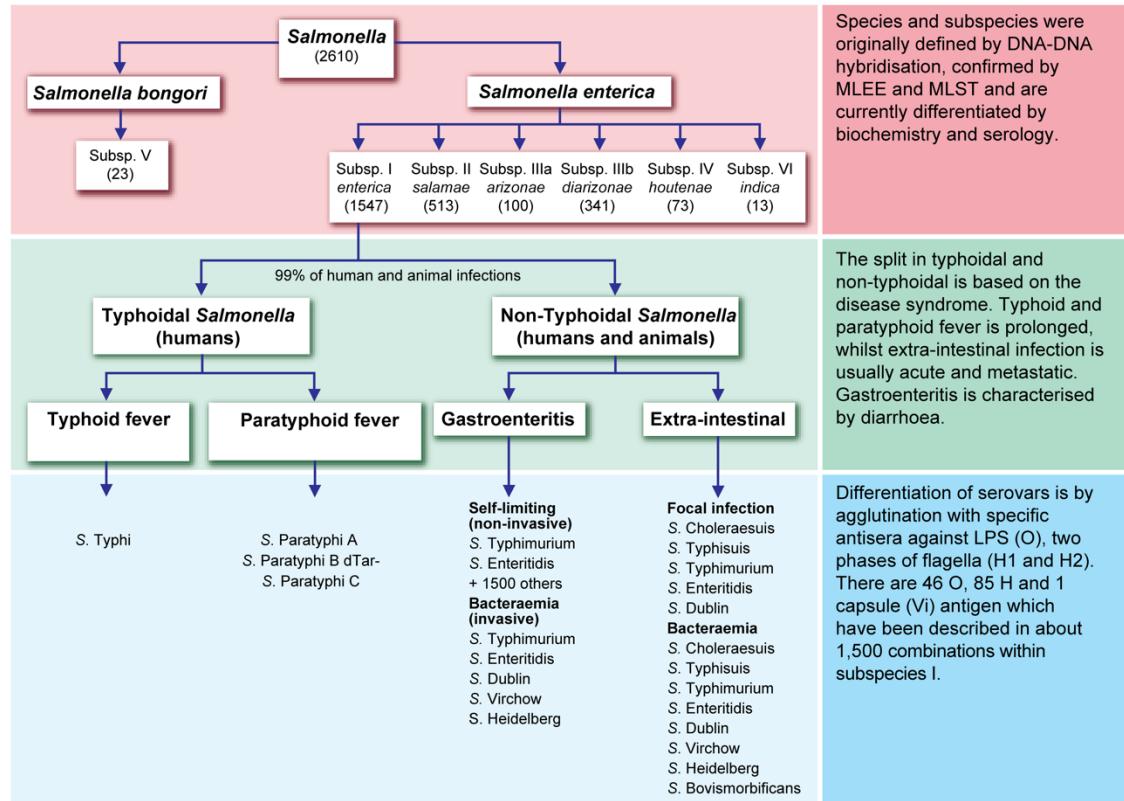
ST-8 Clonal Complex: cc8



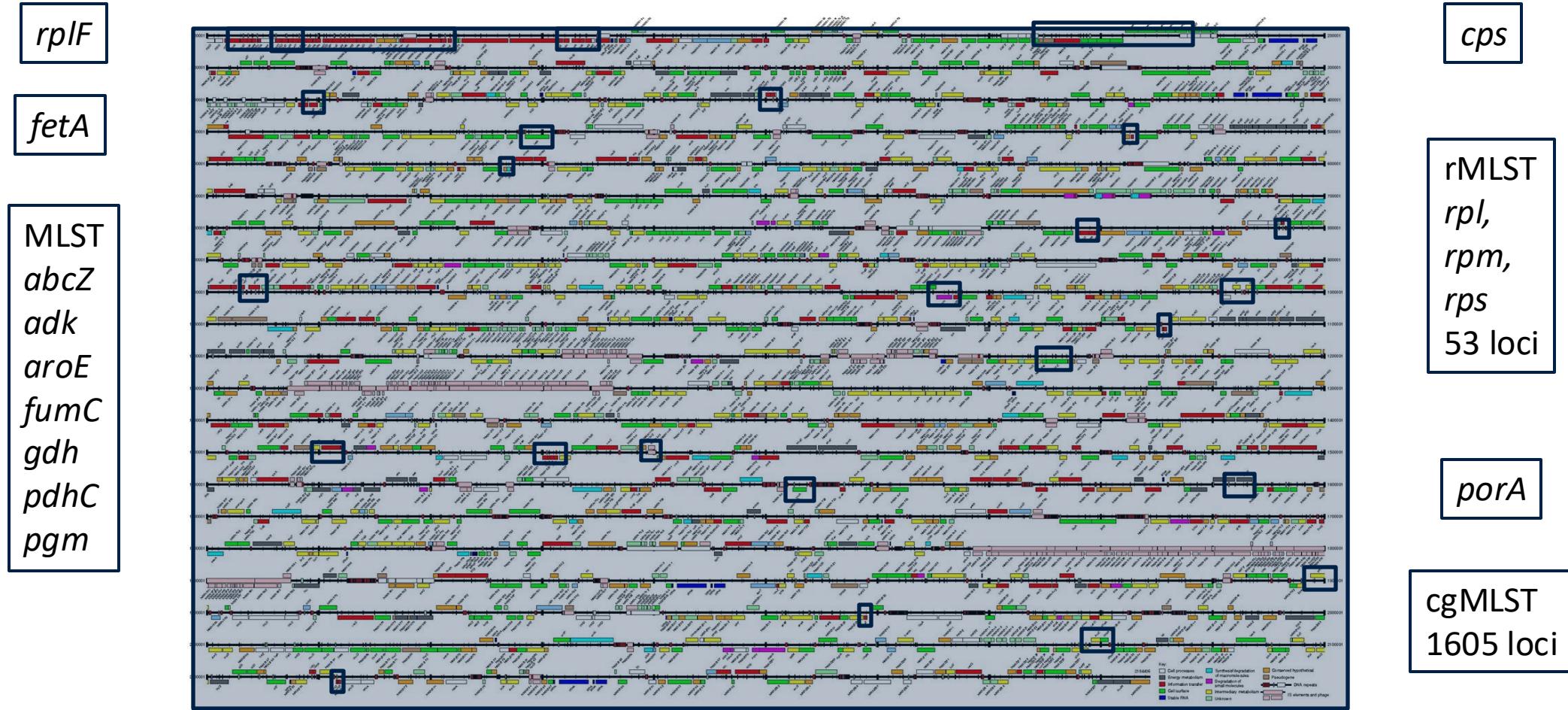
Russell, J. E., Urwin, R., Gray, S. J., Fox, A. J., Feavers, I. M. & Maiden, M. C. (2008). Molecular epidemiology of meningococcal disease in England and Wales 1975-1995, before the introduction of serogroup C conjugate vaccines. *Microbiology* 154, 1170-1177.



# eBURST Groups: *Salmonella*

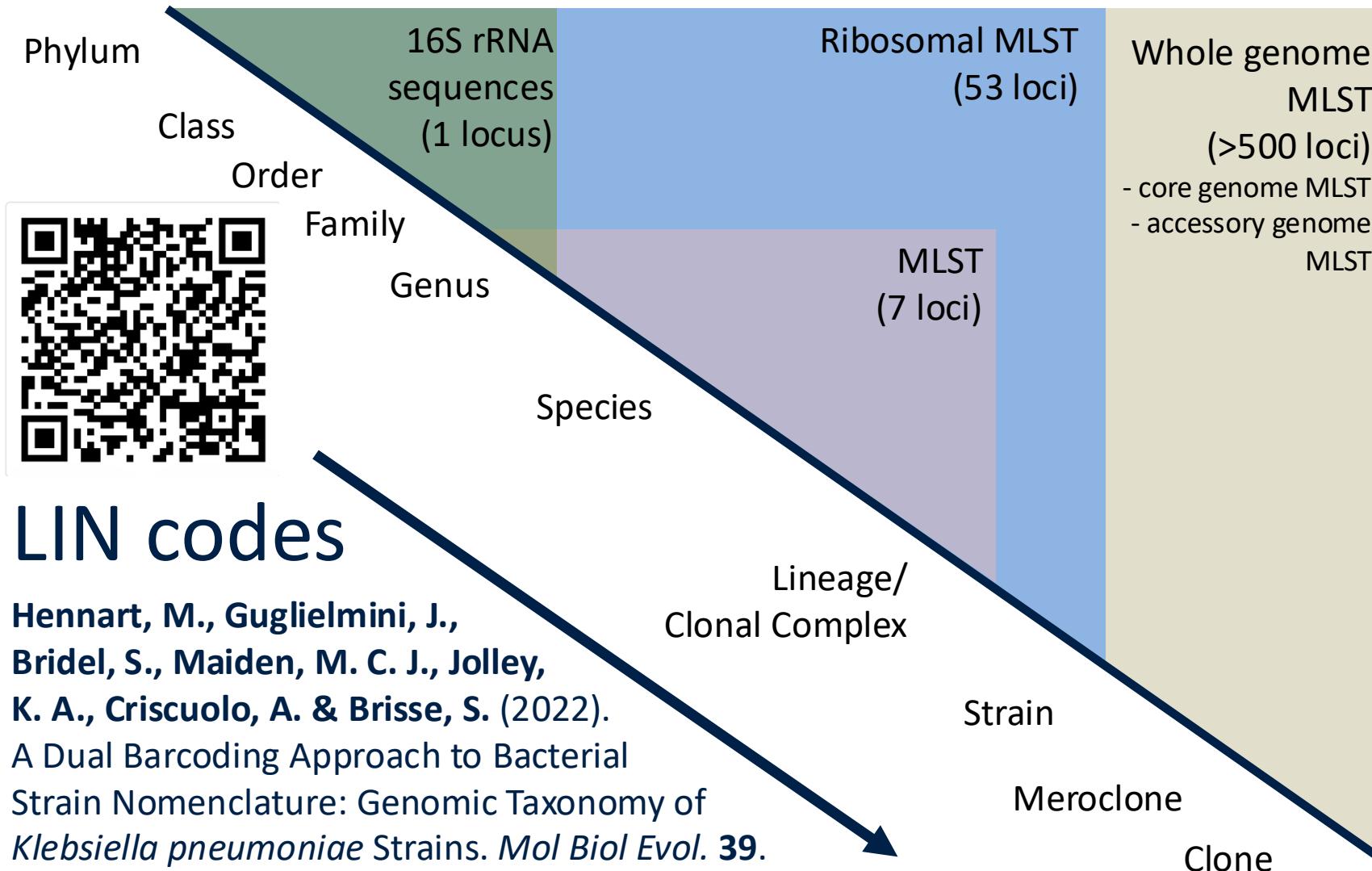


# Whole genome sequences



Parkhill, J., Achtman, M., et al. Spratt, B. G. & Barrell, B. G. (2000). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502-506.  
Bratcher, H. B., Corton, C., Jolley, K. A., Parkhill, J. & Maiden, M. C. (2014). A gene-by-gene population genomics platform: *de novo* assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genomics* **15**, 1138.

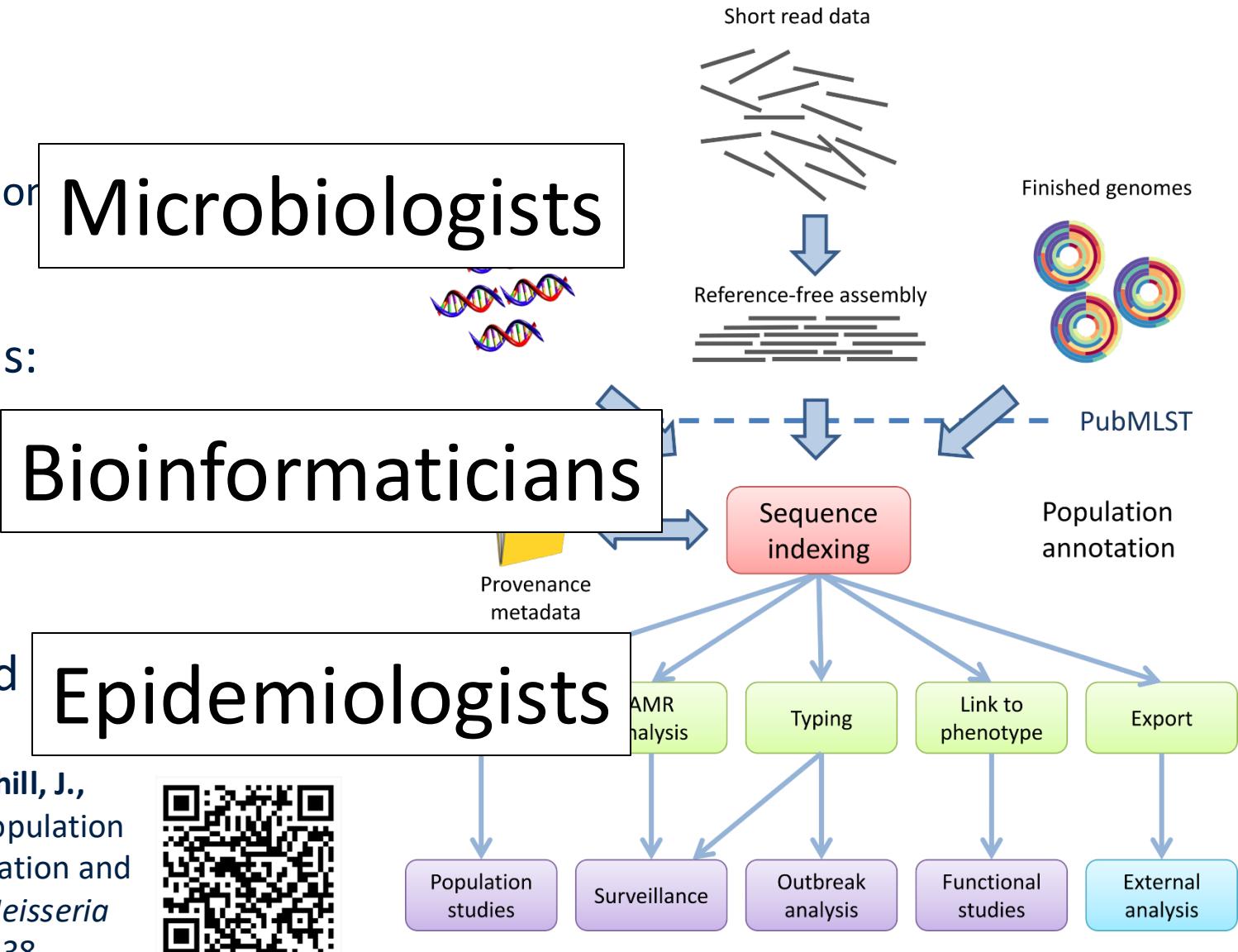
# Conceptual framework: sequence data, nomenclature, phenotype.



Maiden, M. C. J., Jansen van Rensburg, M. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A. & McCarthy, N. D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology* **11**, 728-736.

# PubMLST: data analysis, curation, & dissemination

- Sequence data input:
  - individual gene sequences;
  - *de novo* assembled draft genomes;
  - finished genomes.
- PubMLST (BIGSDB) functions:
  - annotation of loci;
  - linkage to provenance/phenotype (metadata);
  - comparative analyses.



Bratcher, H. B., Corton, C., Jolley, K. A., Parkhill, J., and Maiden, M. C. (2014) A gene-by-gene population genomics platform: *de novo* assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genomics* 15, 1138

A collection of open-access, curated databases that

# Population Genomics and Population Annotation

Identify the population

Sample the population

Sequence  
(*de novo*  
assembly)

Annotate  
gene-by-  
gene

Combine  
data and  
analyse

Disseminate

Organisms search

APPLY

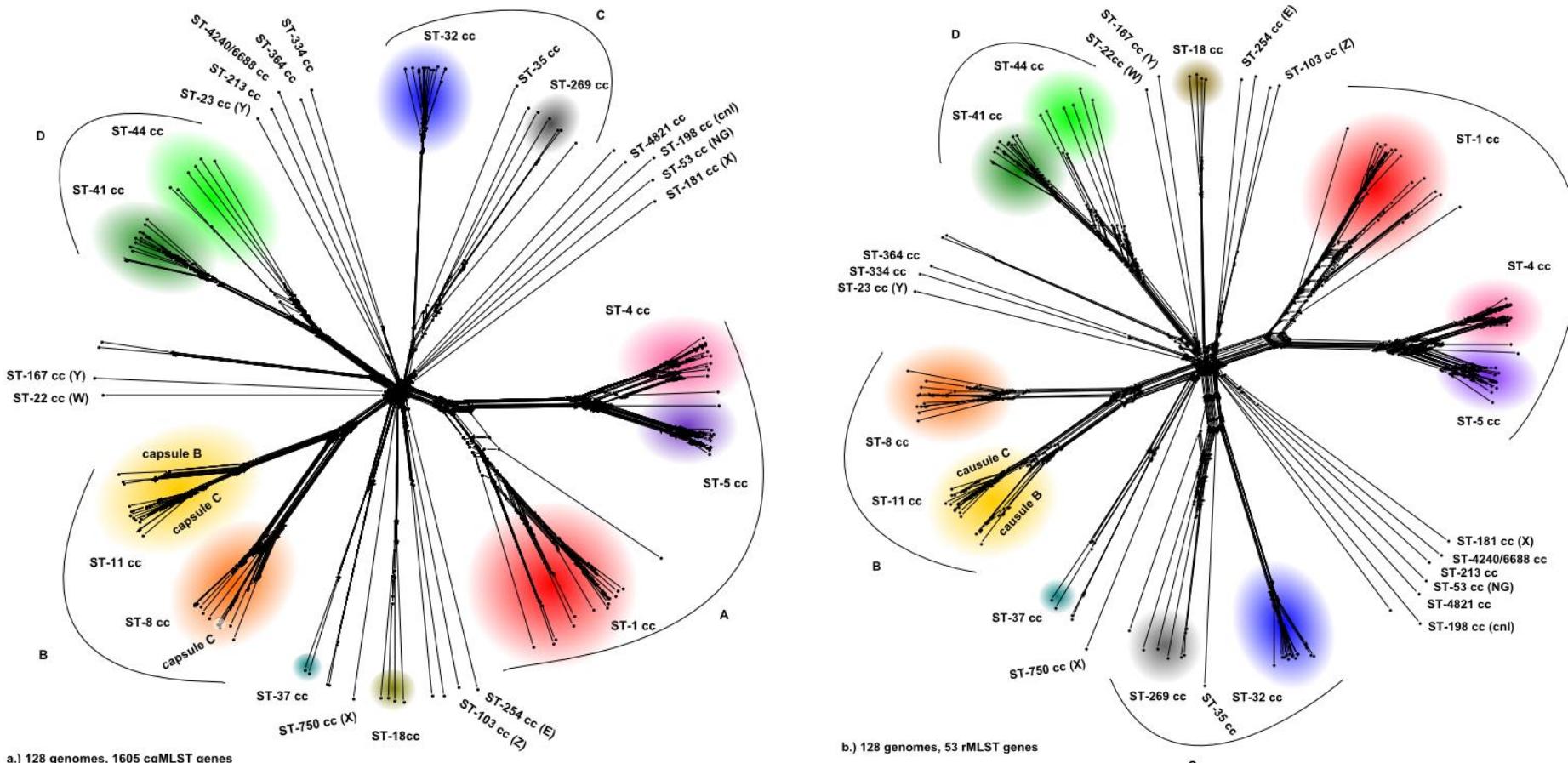
**Jolley, K. A., Bray, J. E. & Maiden, M. C. J. (2018).** Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* **3**, 124.

**Maiden, M. C. J. (2019).** The Impact of Nucleotide Sequence Analysis on Meningococcal Vaccine Development and Assessment. *Front Immunol* **9**, 3151.

isolate records.

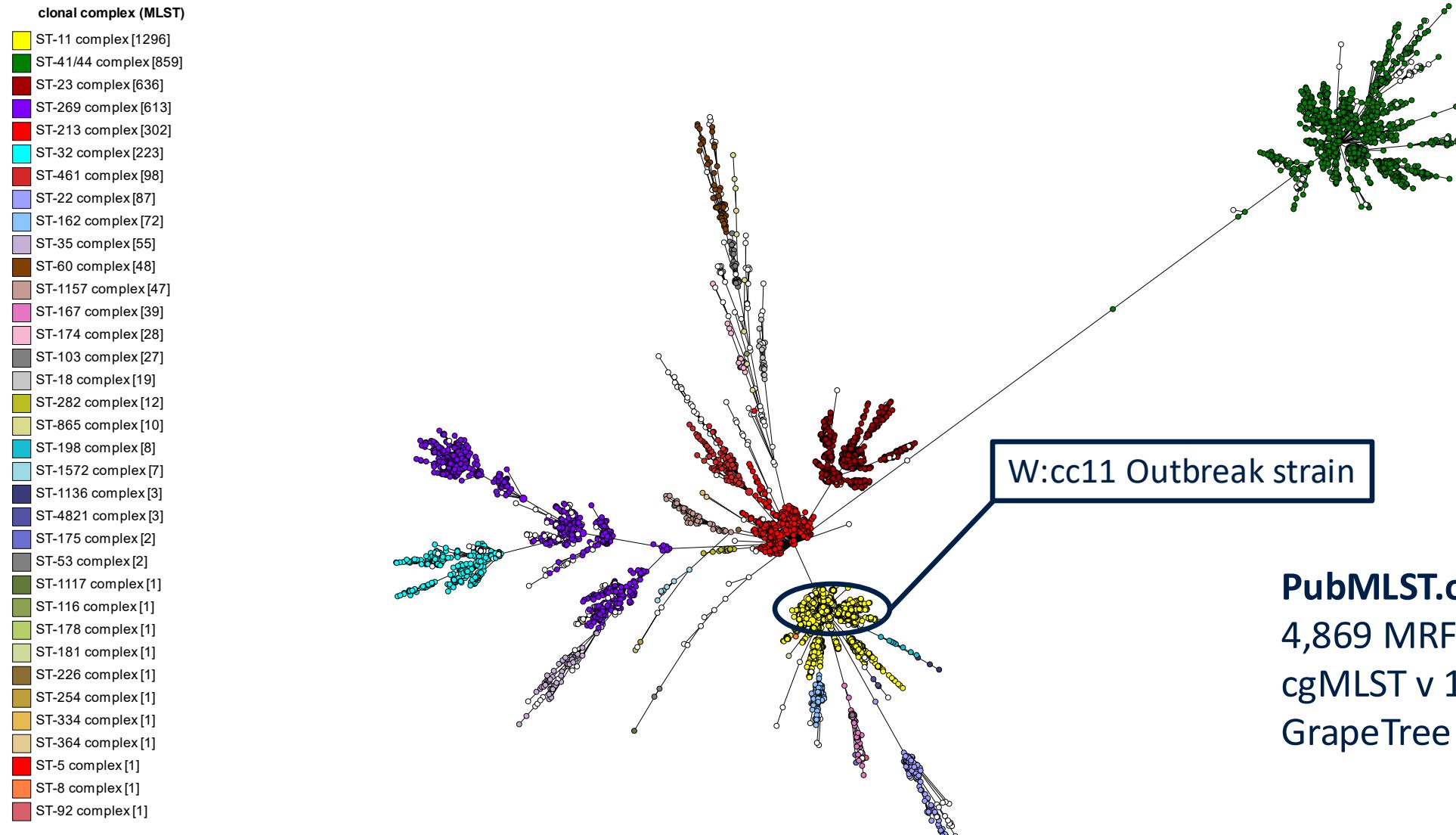
profiles, or isolate records. Isolates may be accompanied by a genome assembly.

# Meningococcal genealogies: cgMLST, rMLST, & MLST



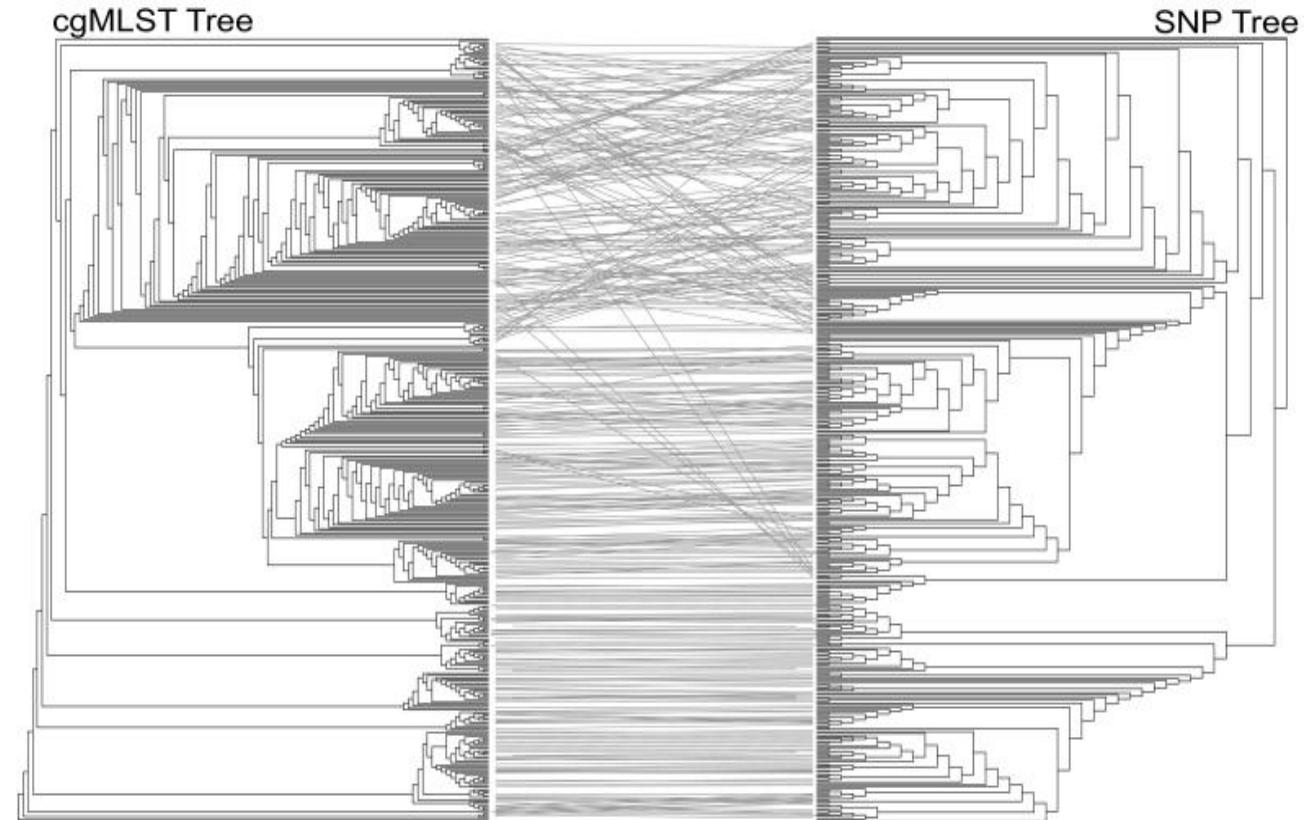
**Bratcher, H. B., Corton, C., Jolley, K. A., Parkhill, J. & Maiden, M. C. (2014). A gene-by-gene population genomics platform: *de novo* assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. BMC Genomics 15, 1138.**

# Meningitis Research Foundation Meningococcal Genome Library 2010-2020



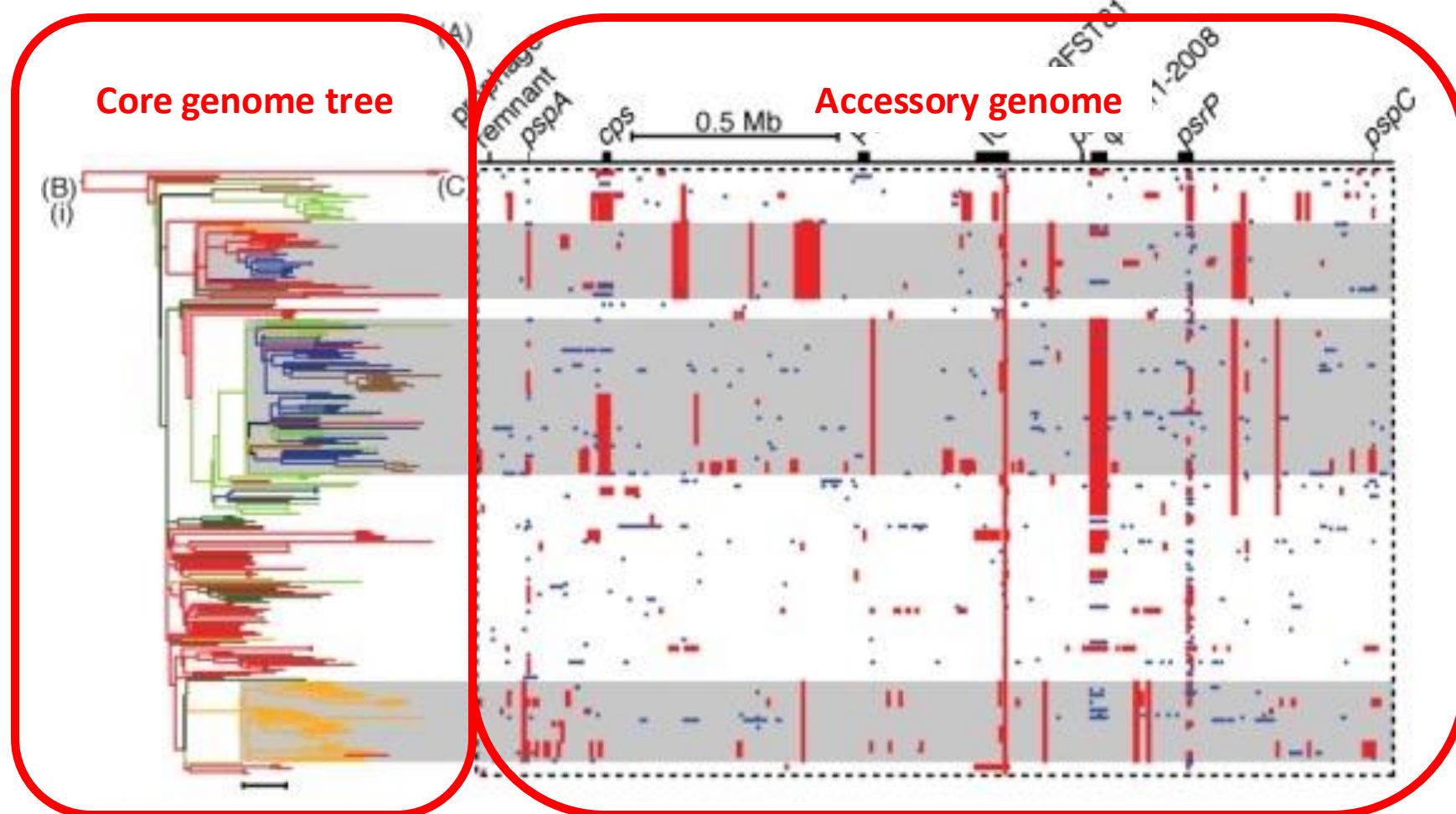
# cgMLST vs 'core SNPs'

- Essentially based on the same data.
- cgMLST can be used to recover 'core SNPs'.
- cgMLST provides a stable nomenclature that is reference free.



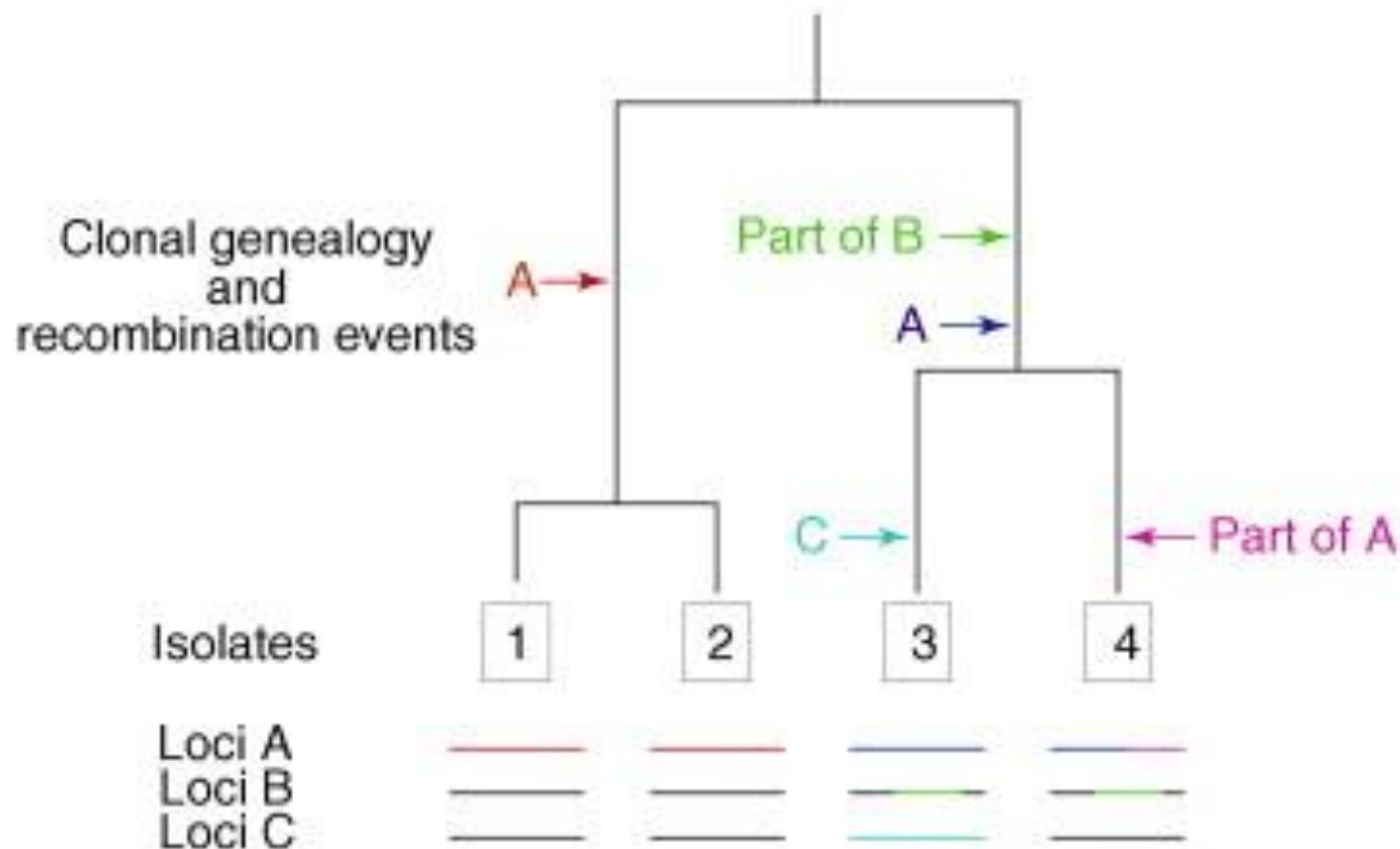
Pearce, M. E., Alikhan, N. F., Dallman, T. J., Zhou, Z., Grant, K. & Maiden, M. C. J. (2018). Comparative analysis of core genome MLST and SNP typing within a European *Salmonella* serovar Enteritidis outbreak. *International Journal of Food Microbiology* 274, 1-11.

# Gubbins: an analysis of PMEN1 clone



Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., Parkhill, J. & Harris, S. R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research* 43, e15.

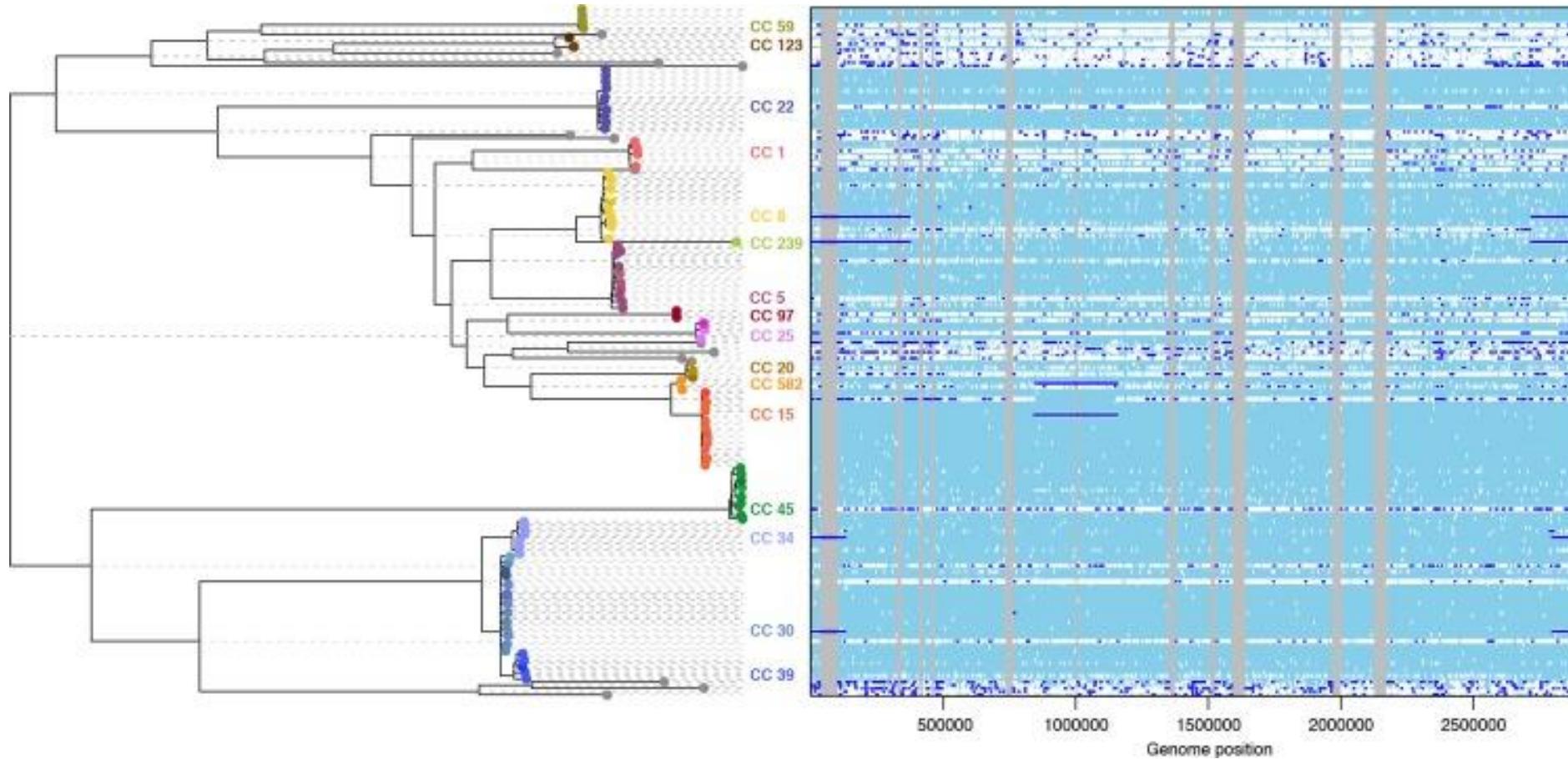
# The ClonalFrame Model



TRENDS in Microbiology

Didelot, X. & Maiden, M. C. (2010). Impact of recombination on bacterial evolution. *Trends in Microbiology* 18, 315-322.

# ClonalFrameML: recombination in *S. aureus*



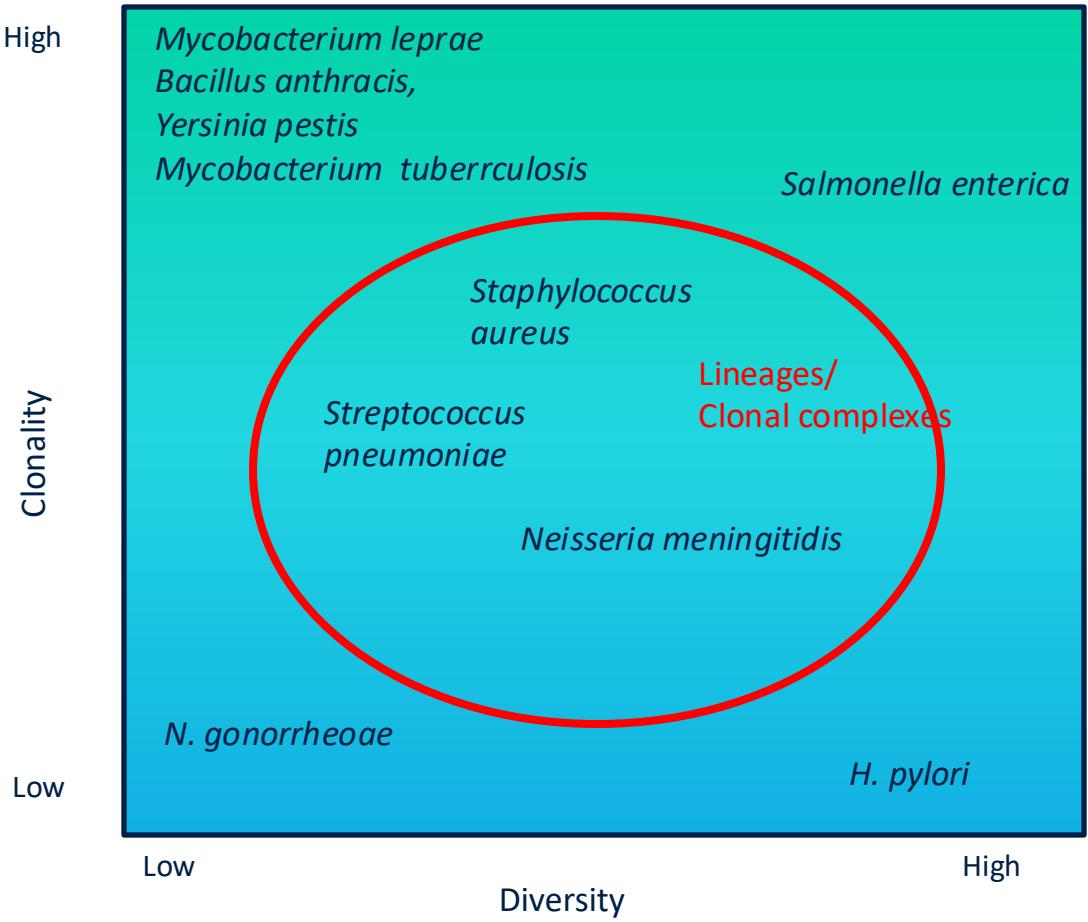
**Didelot, X. & Wilson, D. J. (2015).** ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLoS Computational Biology* 11, e1004041.

# Population structure and diversity

In conclusion:

- different levels of clonal signal are observed in different bacterial populations;
- this is a consequence of differing relative rates of recombination to mutation;
- however, other forces will also play a role.

**Didelot, X. & Maiden, M. C. J. (2010).**  
Impact of recombination on bacterial evolution. *Trends Microbiol* 18, 315-322.



# Take home messages

1. The clinical question is all important.
  - What is the end point that you wish to achieve?
2. Pathogen DNA is a very useful source of information.
  - What are the most appropriate data and technology?
3. Bacteria are very highly diverse.
  - Don't expect that 'one size fits all' – analyses and interpretations will differ for different bacteria (sometimes even closely related ones).
4. The impact of HGT varies among bacteria.
  - This affects typing and interpretation: how clonal is my pathogen?

