A Formal Model of Clonal Expansion in (Bacterial) Population Genetics

Alice Ledda and Xavier Didelot

DIDE - Imperial College London

MCEB, June 13 2016

Who infected whom?

Who infected whom?

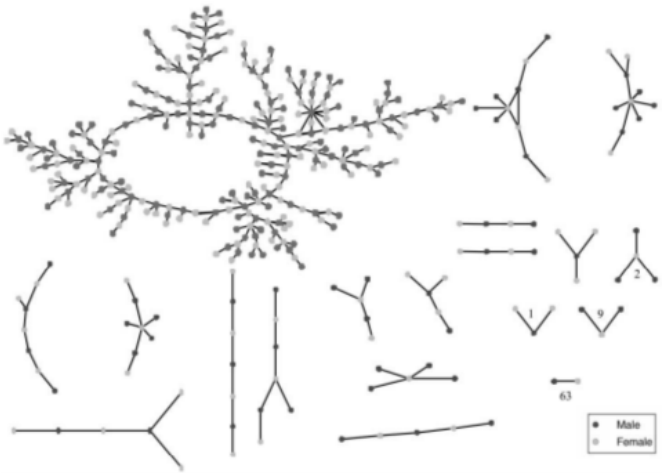


FIG. 2.—The direct relationship structure at Jefferson High

Phylogenetic tree describing the evolutionary relationships of full-length porB gene sequences of the high-level ceftriaxone-resistant Neisseria gonorrhoeae strain H041 compared with those of previously published N. gonorrhoeae penA mosaic isolates (11).

Makoto Ohnishi et al. Antimicrob. Agents Chemother. 2011;55:3538-3545

Antimicrobial Agents and Chemotherapy

Staphylococcus aureus

- Gram-positive bacteria
- usually found in nose and skin
- usually not harmful

Staphylococcus aureus

- Gram-positive bacteria
- usually found in nose and skin
- usually not harmful

BUT

it is the most frequent cause of hospital bacteremia

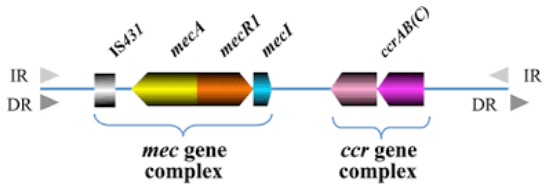
used as a proxy of poor infection control

MRSA

Meticillin Resistant Staphylococcus Aureus

resistance to a class of antimicrobials
the β - lactams (*pennicillins*, *cefalosporins*, *carapenems*)

All in a single genetic element!

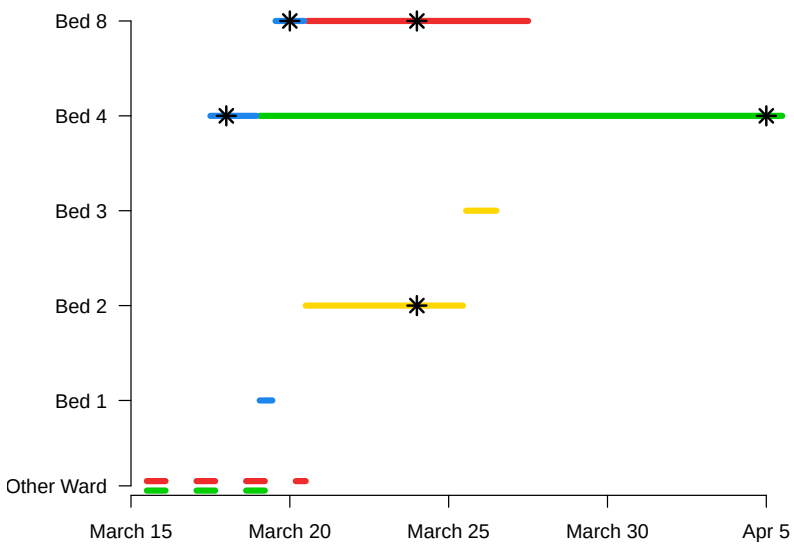


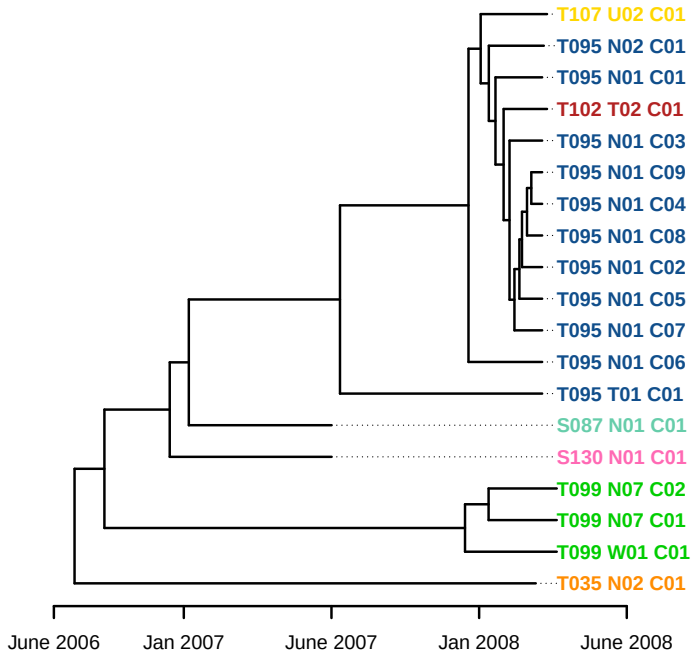
the SCCmec cassette

First seen 6 months after the introduction of Meticillin
At the beginning just present in few hospital strains.
Now present everywhere.

Genetic Epidemiology of MRSA Staph: an example

Samples from patients (and some staff) in an ICU collected and tested sistematically everyday for 3 months

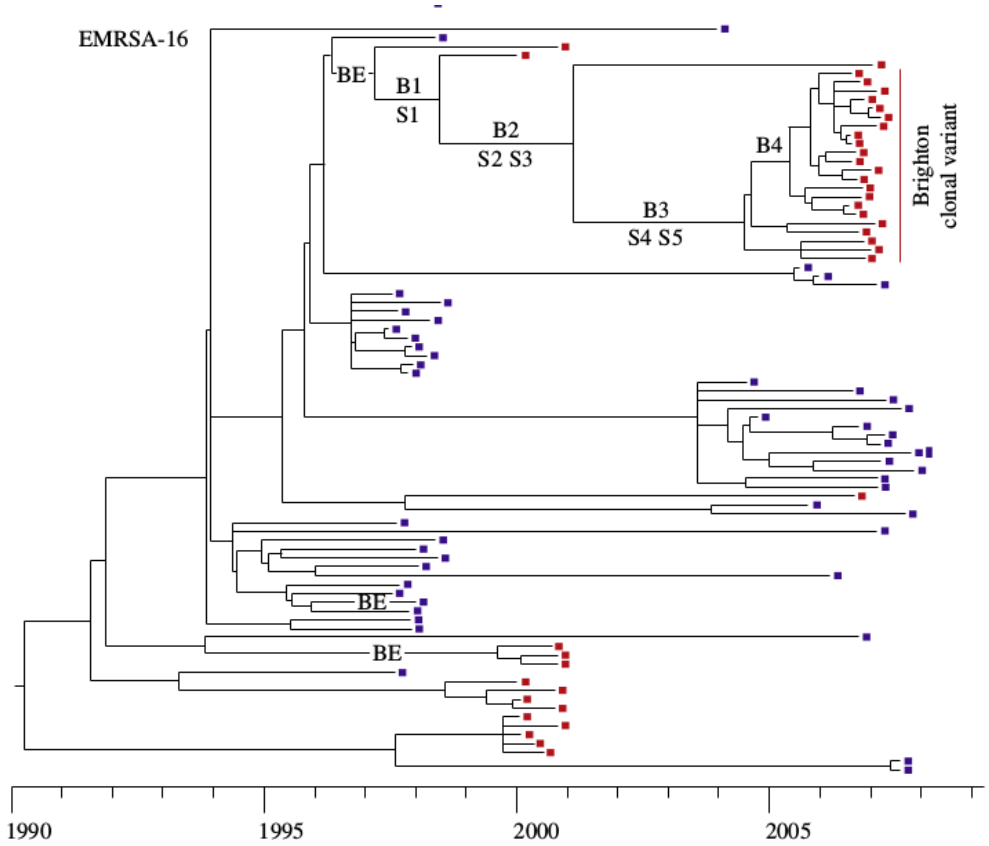




Ledda et al, in preparation

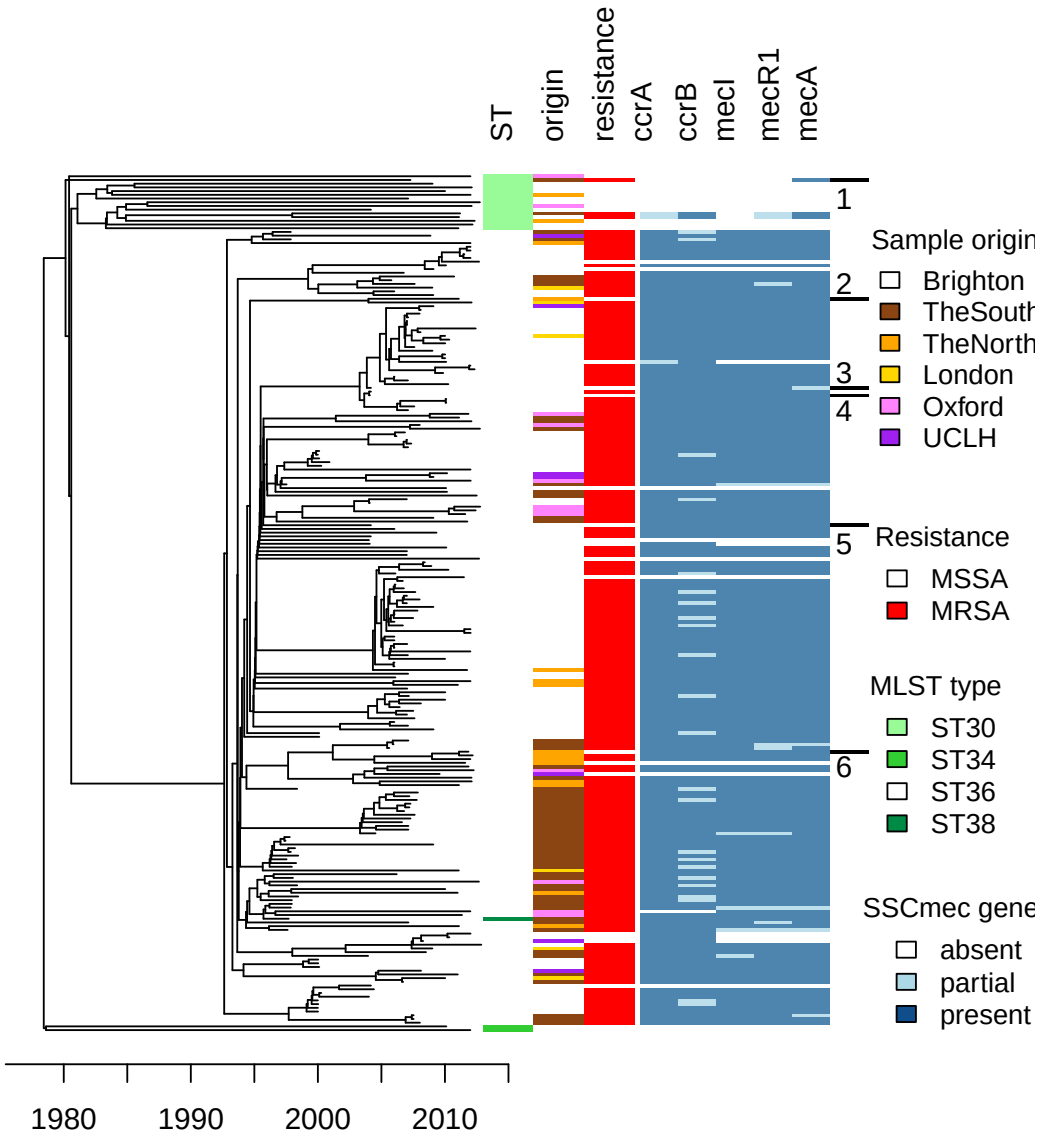
Genetic Epidemiology of MRSA Staph: another example

Samples from an outbreak in Brighton in 2007



Miller et al. 2013

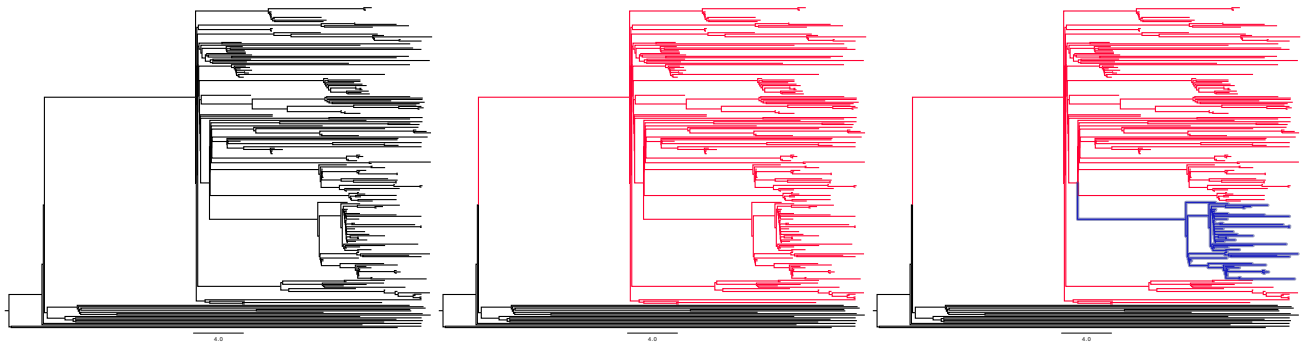
Genetic Epidemiology of MRSA Staph: another example



Aim of the work

Can we produce a tool that given a phylogenetic tree tells us:

- if there is an outbreak
- where is the outbreak (on the tree)
- when it started
- how strong it is



so that we can trigger public health responses

Outbreak ⇒ Clonal Expansion?

Outbreak as a medical expression is not necessarily linked to population genetics.

Clonal expansion: a subpopulation has recently become abundant and widespread.

Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4384–4388, May 1993
Population Biology

How clonal are bacteria?

(recombination/linkage disequilibrium/population structure/genetic transformation/parasite evolution)

JOHN MAYNARD SMITH, NOEL H. SMITH, MARIA O'ROURKE, AND BRIAN G. SPRATT*

School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom

Contributed by John Maynard Smith, January 25, 1993

ABSTRACT **Data from multilocus enzyme electrophoresis of bacterial populations were analyzed using a statistical test designed to detect associations between genes at different loci. Some species (e.g., *Salmonella*) were found to be clonal at all levels of analysis. At the other extreme, *Neisseria gonorrhoeae* is panmictic, with random association between loci. Two intermediate types of population structure were also found. *Neisseria meningitidis* displays what we have called an "epidemic" structure. There is significant association between loci, but this arises only because of the recent, explosive, increase in**

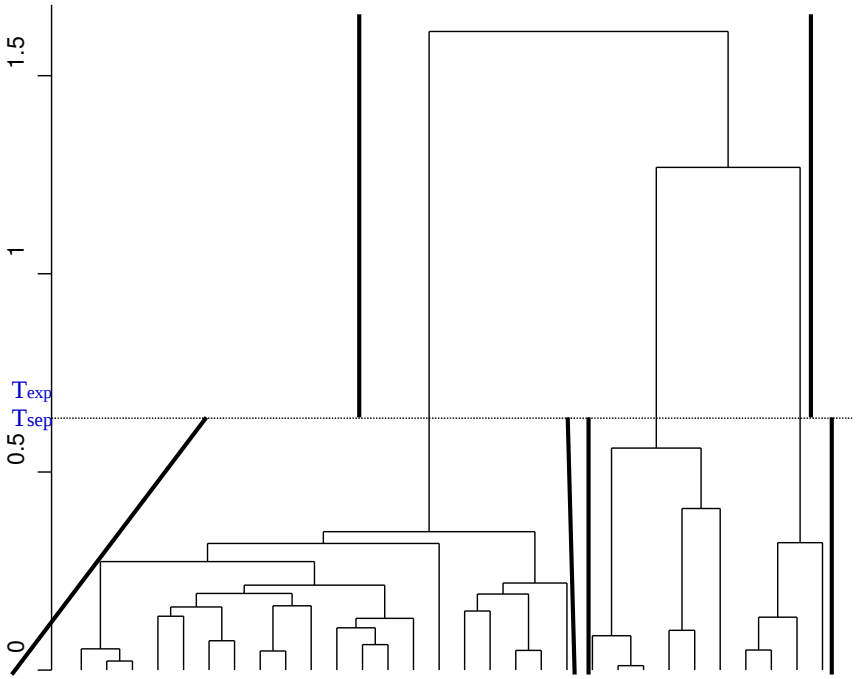
That *E. coli* populations are clonal has been elevated, with notable exceptions (16, 17, 22), to the status of a paradigm extending to all bacterial populations. The data supporting clonality in other species of bacteria tend to rely heavily upon the demonstration of high coefficients of linkage disequilibrium by MLEE studies and the frequent recovery of one or a few multilocus genotypes. However, linkage disequilibrium can arise in bacterial populations in which recombination is frequent, in several ways:

(*i*) The samples analyzed may consist of a mixture of

Why now?

- MLST could not detect the different clonal subpopulations
- Whole genome sequencing can
- More and more full genome sequences
- More outbreaks of MRSA *Staph.* in the last 20 years
- More resistance

Our approach



T_{sep}: time at which the two population split - no migration

T_{exp}: time at which the expanding population starts to expand

α: expansion rate

Our approach

Taking into account this model we developed a
Monte Carlo Markov Chain
that infers

- the expansion rate α
- the clade where the expansion starts
- the ratio between the two subpopulations (assuming/non assuming fair sampling)

Likelihood:
Non expanding population

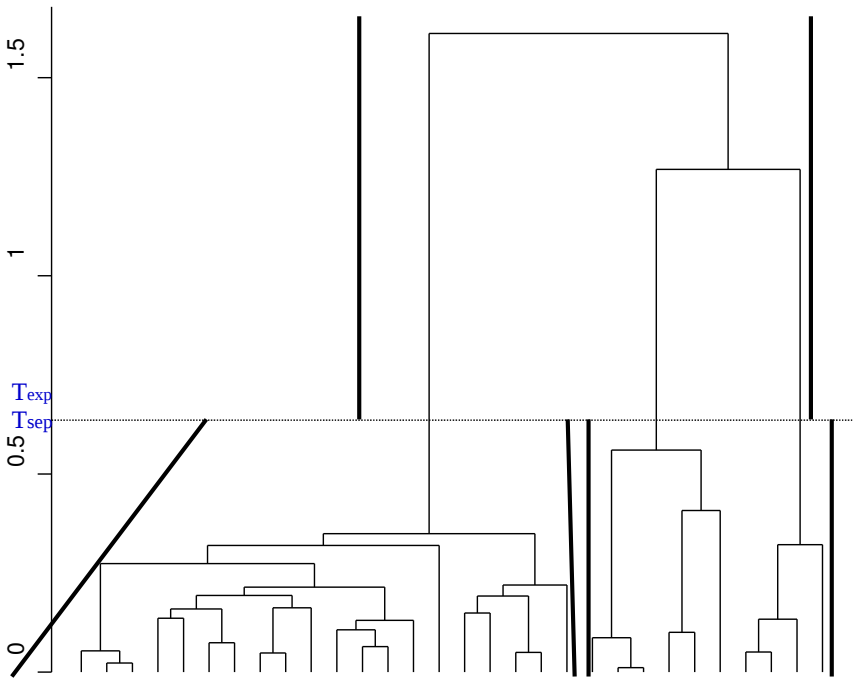
$$f_G(g|\theta) = \frac{1}{\theta^{n-1}} \prod_{i=2}^{2n-1} \exp^{(-k_i(k_i-1)/2\theta)(t_i - t_{i-1})}$$

Population expanding at a rate α

$$f_G(g|\theta,\alpha) = \frac{1}{\theta^{n-1}} \prod_{i=2}^{2n-1} \exp^{\alpha t_i} \exp^{(-k_i(k_i-1)/2\theta r)(\exp^{\alpha t_i} - \exp^{\alpha t_{i-1}})}$$

Drummond et al. 2002

Bounded Coalescent



All the expanding supopulation has to coalesce before it merges into the other.

$\Downarrow$

We have a boundary condition for the coalescent times

$$T^{sub}_{MRCA} < T_{sep}$$

$\Downarrow$

Bounded coalescent

Rasmussen and Kellis 2012

Bounded Coalescent

Formula 1 in Rasmussen and Kellis should be

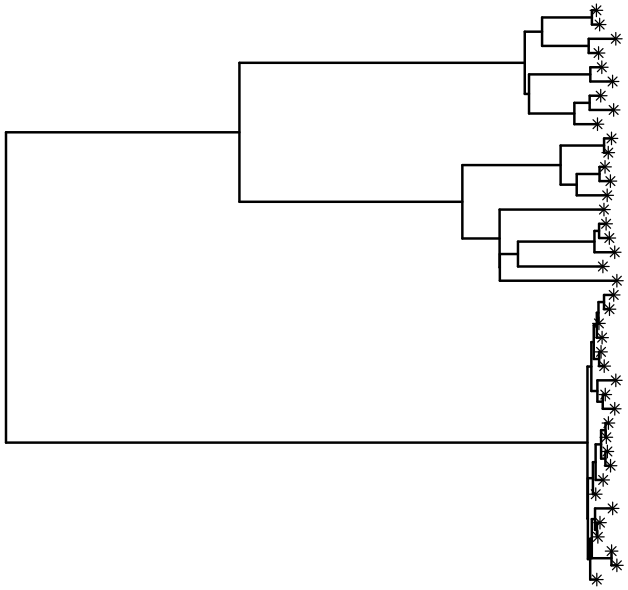
$$
\begin{aligned}
P(t_2, \ldots, t_k | t_{MRCA} < t^*, k, N) &= \frac{\prod_{j=2}^{k} P(t_j | j, N)}{P(T_{MRCA} < \sum_{i=2}^{k} t_i | k, N)} \\
&= \prod_{j=2}^{k} \frac{P(t_j | j, N) P(T_{MRCA} < t^* - \sum_{i=j}^{k} t_i | j-1, N)}{P(T_{MRCA} < \sum_{i=j+1}^{k} t_i | j, N)}
\end{aligned}
$$

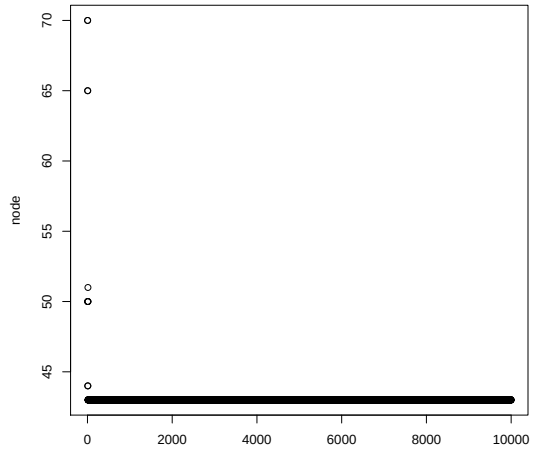
where

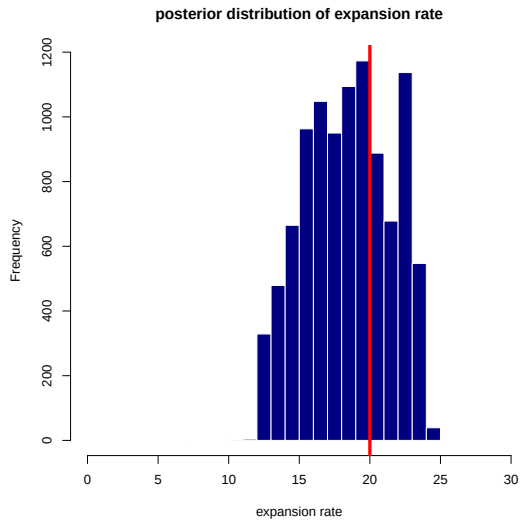
$$
\begin{aligned}
P(T_{MRCA} < t^* - \sum_{i=j}^{k} t_i | j-1, N) &= \int_0^{t^* - \sum_{i=j}^{k} t_i} f_{MRCA}(t) dt \\
&= \sum_{i=2}^{k} \left(1 - e^{-\frac{i(i-1)(t^* - \sum_{i=j}^{k} t_i)}{2}} \right) \prod_{j=2; j \neq i}^{k} \frac{j(j-1)}{i(i-1) - j(j-1)}
\end{aligned}
$$

Tavare 1984

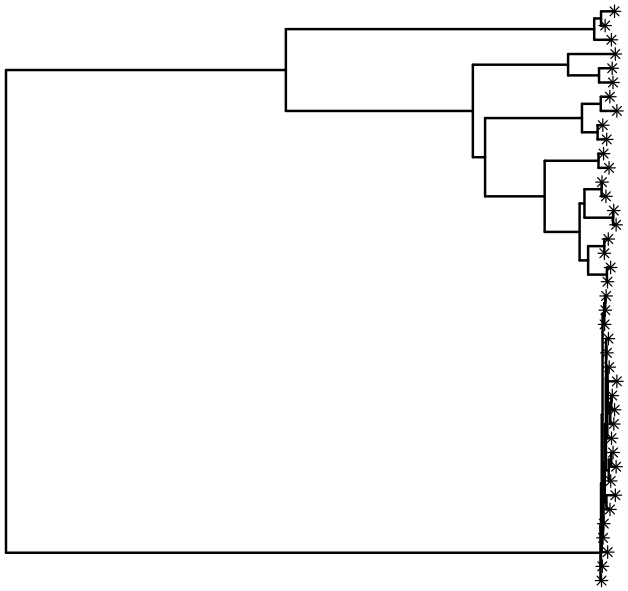
Example 1: expansion rate 20

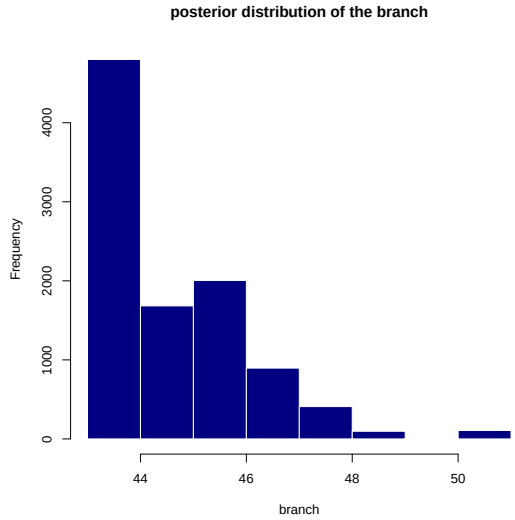


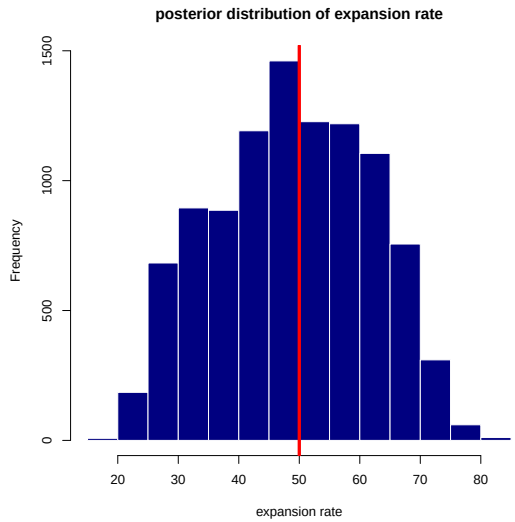




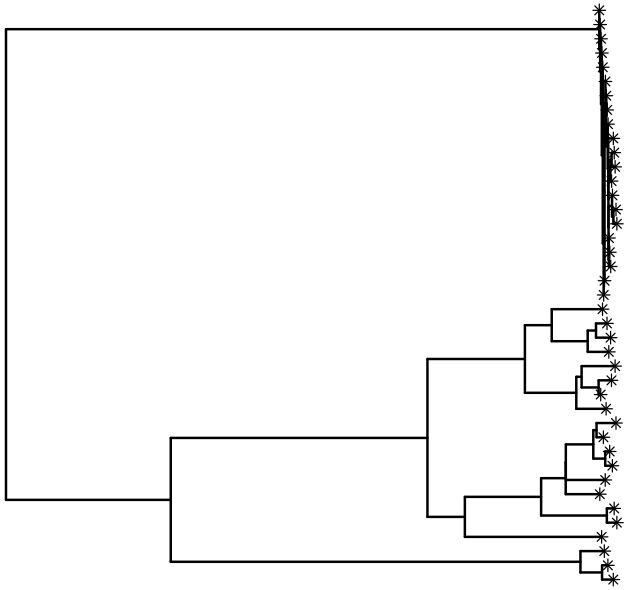
Example 2: expansion rate 50

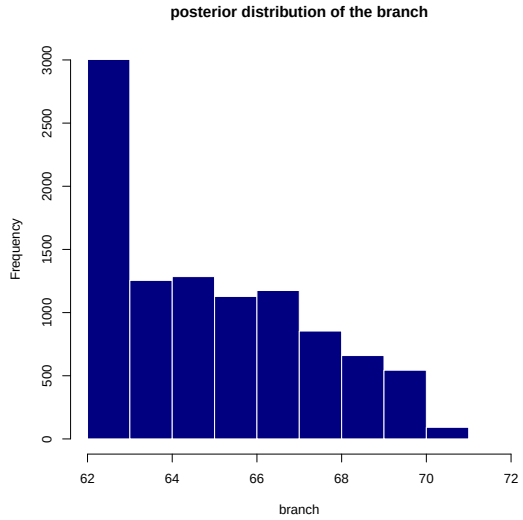


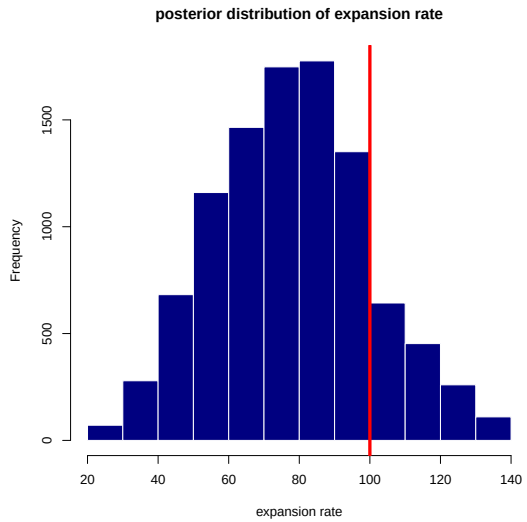




Example 2: expansion rate 100







Ongoing and future work

- extended to non-ultrametric trees
- it is *pathogen-indipendent*
- relax the fair sampling hypothesis
- infer times, not clades
- infer multiple expansions at different expansion rates
- relax the fair sampling hypothesis
- Approaches to monitor clonal expansion in (almost) real time
- Understand the molecular bases of observed clonal expansion
- Relations between extent of antibiotic resistance and rate of clonal expansion

Acknowledgements

Xavier Didelot (Imperial College London)

Luca Ferretti (Pirbright Institute)

Thank you for listening!