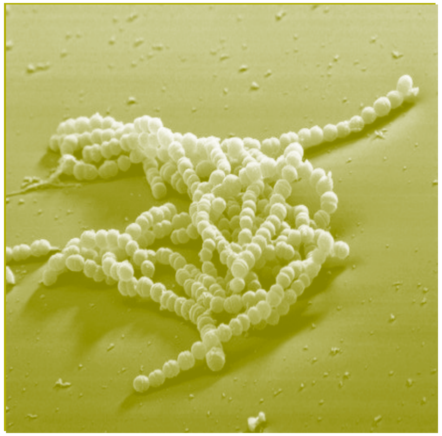
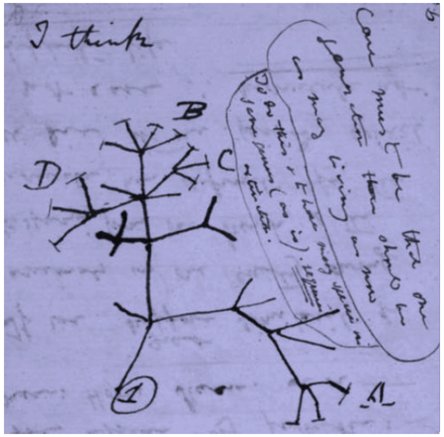
+

Detection of recombination events in bacterial genomes





THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Dr Nouri BEN ZAKOUR
SCMB – AID – University of Queensland

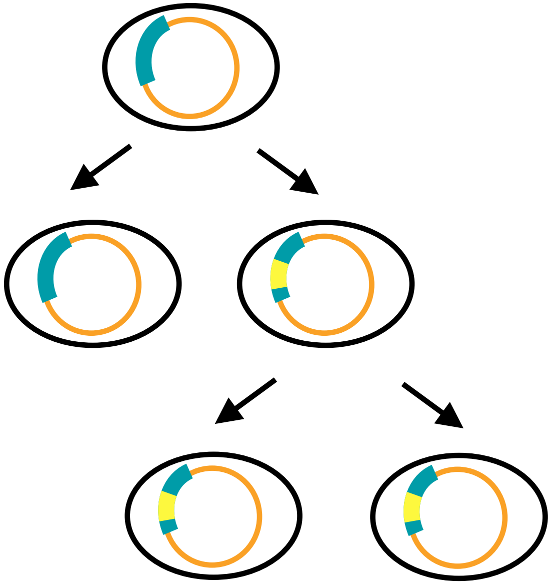
Winter School – 8th of July 2014

Outline

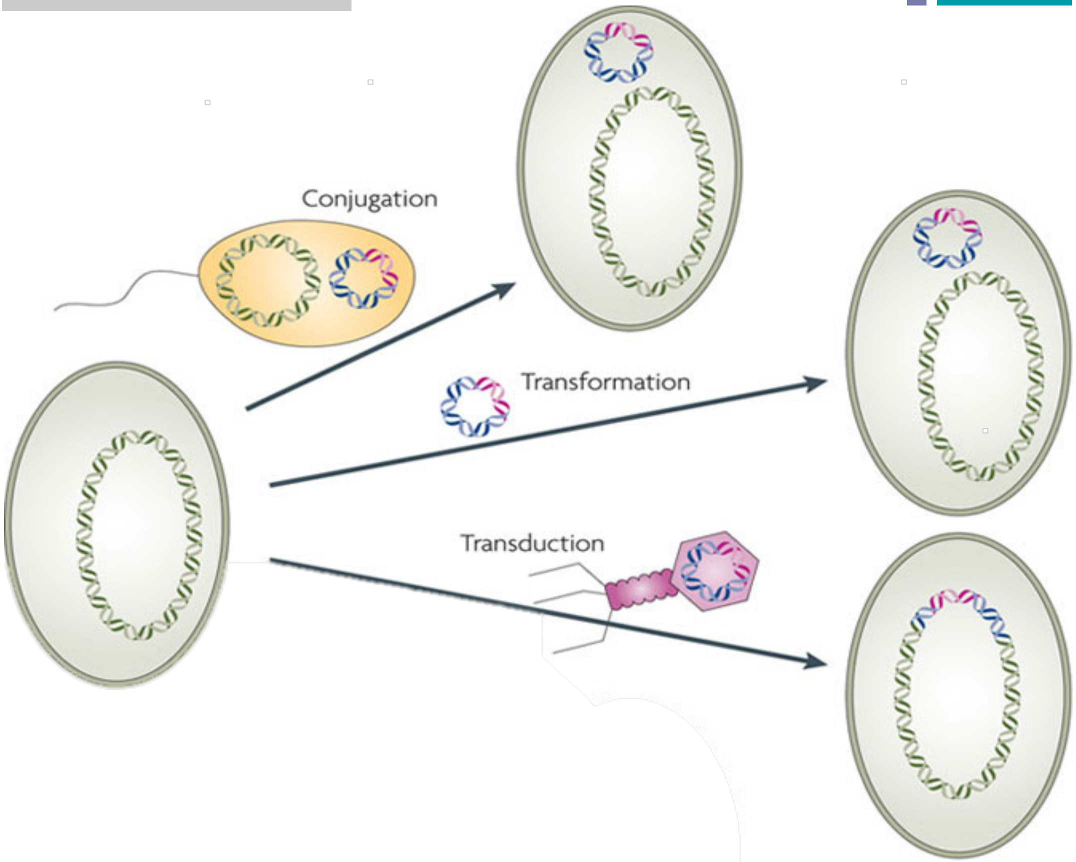
- Concepts
  - Recombination and evolution
- Methods currently used to detect recombinant regions
  - Principle and challenges
  - Approaches
- Case study:
  - Superbugs and recombination

Bacterial evolution

vertical



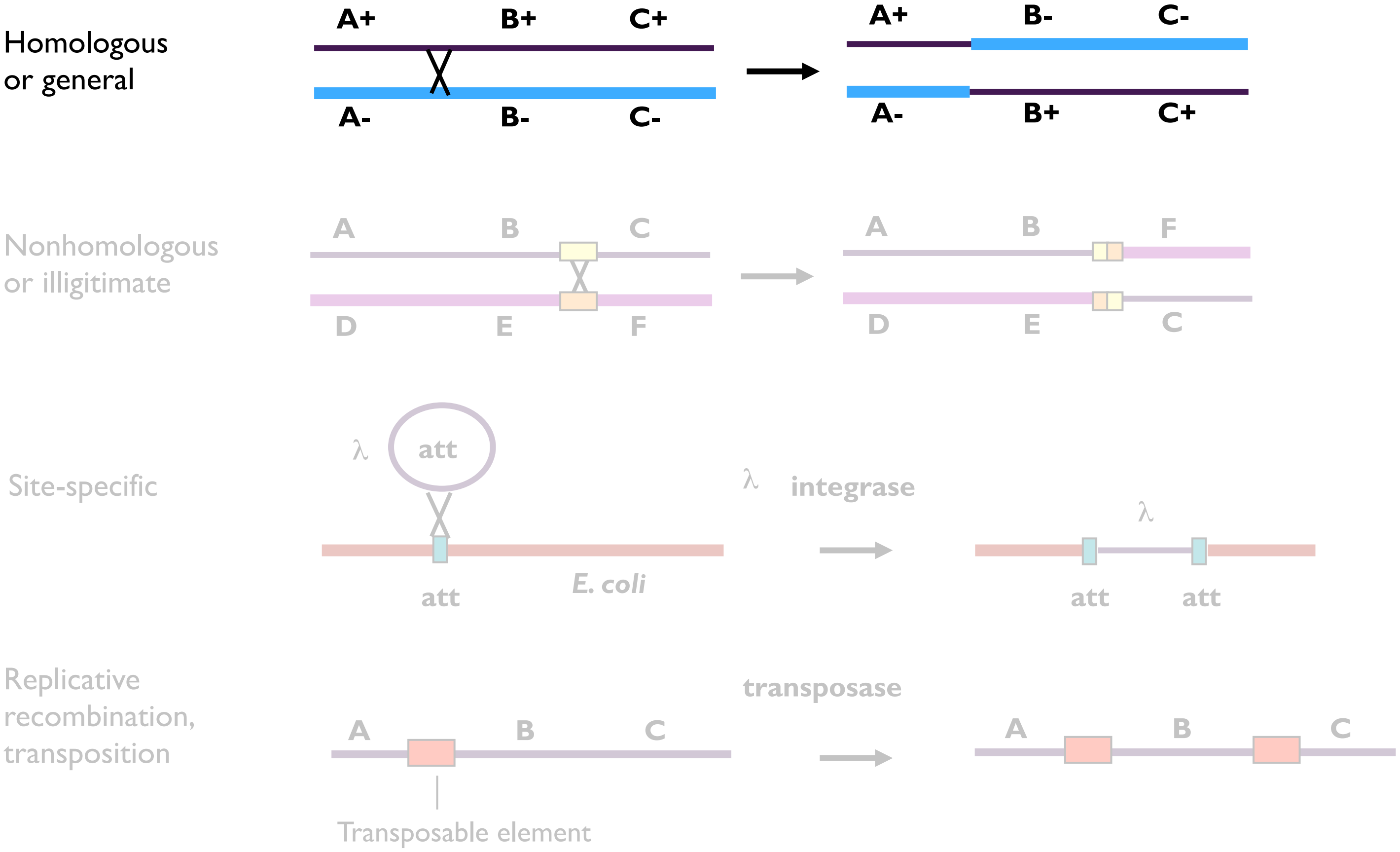
horizontal



Bacteria can evolve not only by accumulating point mutations, but also by acquiring foreign DNA via genetic recombination.

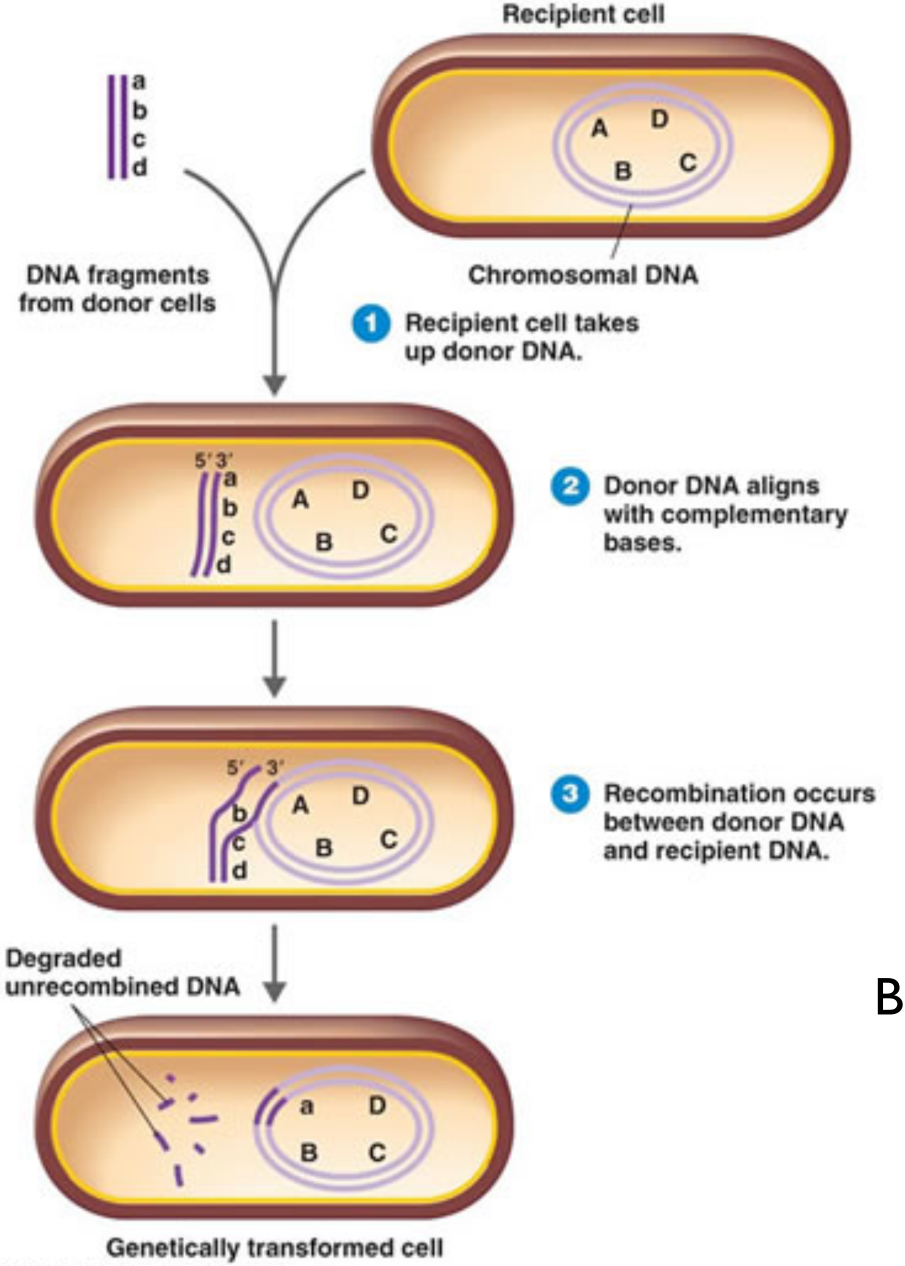
Definition of recombination

- Breaking and rejoining of two parental DNA molecules to produce new DNA molecules



Adapted from Ross Hardison, 2011

Recombination of homologous chromosomal regions

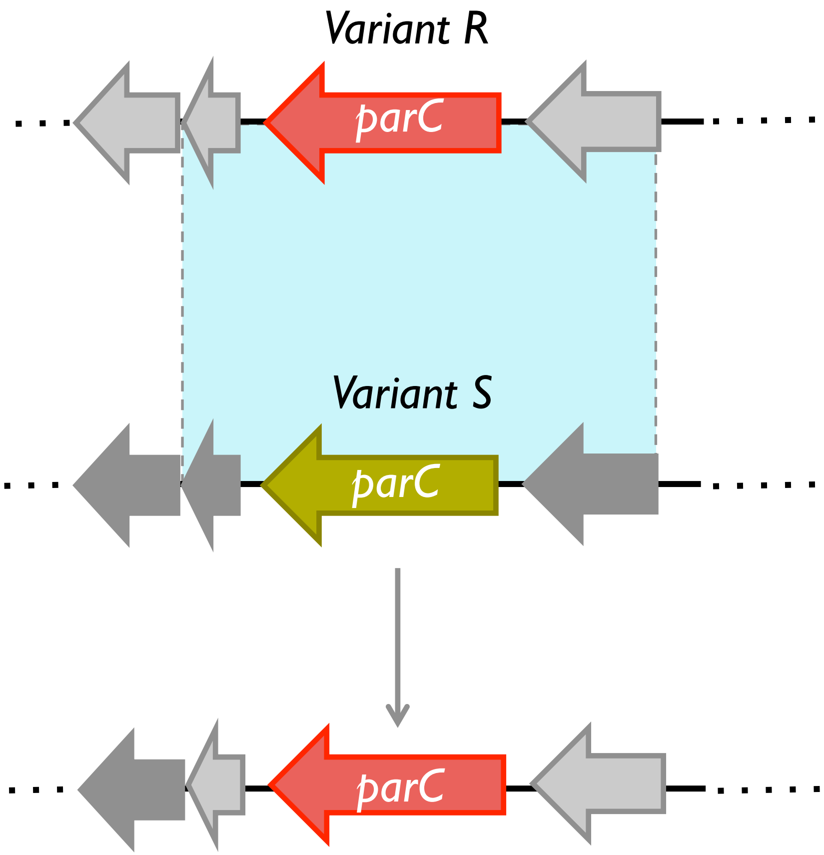


Bacteria can “reshuffle” alleles present in a bacterial population through this mechanism, which allows them to exchange homologous DNA regions.

Functional impacts of recombination on bacteria

Recombination can mediate large evolutionary jumps in bacterial genomes by rapidly spreading variants associated with increased:

- virulence (surface proteins, antigenic variation, etc.)
- fitness (carbohydrates metabolism, etc.)
- antibiotic resistance



Strain A
Resistant to
Fluoroquinolones (FQ)

RECOMBINATION

Strain B
Sensitive to FQ

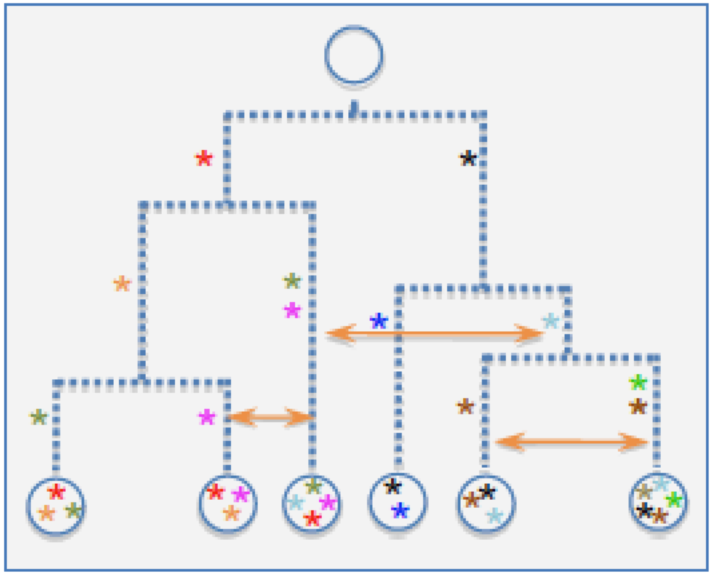
Resistant to FQ

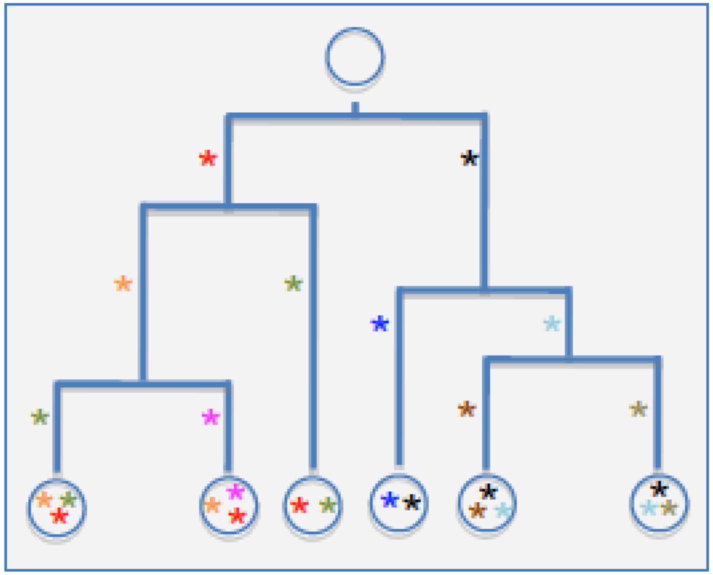
= ONE SINGLE EVENT

Recombination is variable among bacterial species

Non-clonal

Clonal





Helicobacter pylori

Mycobacterium tuberculosis

Polymorphic
Free living
Naturally transformable
High rate of recombination

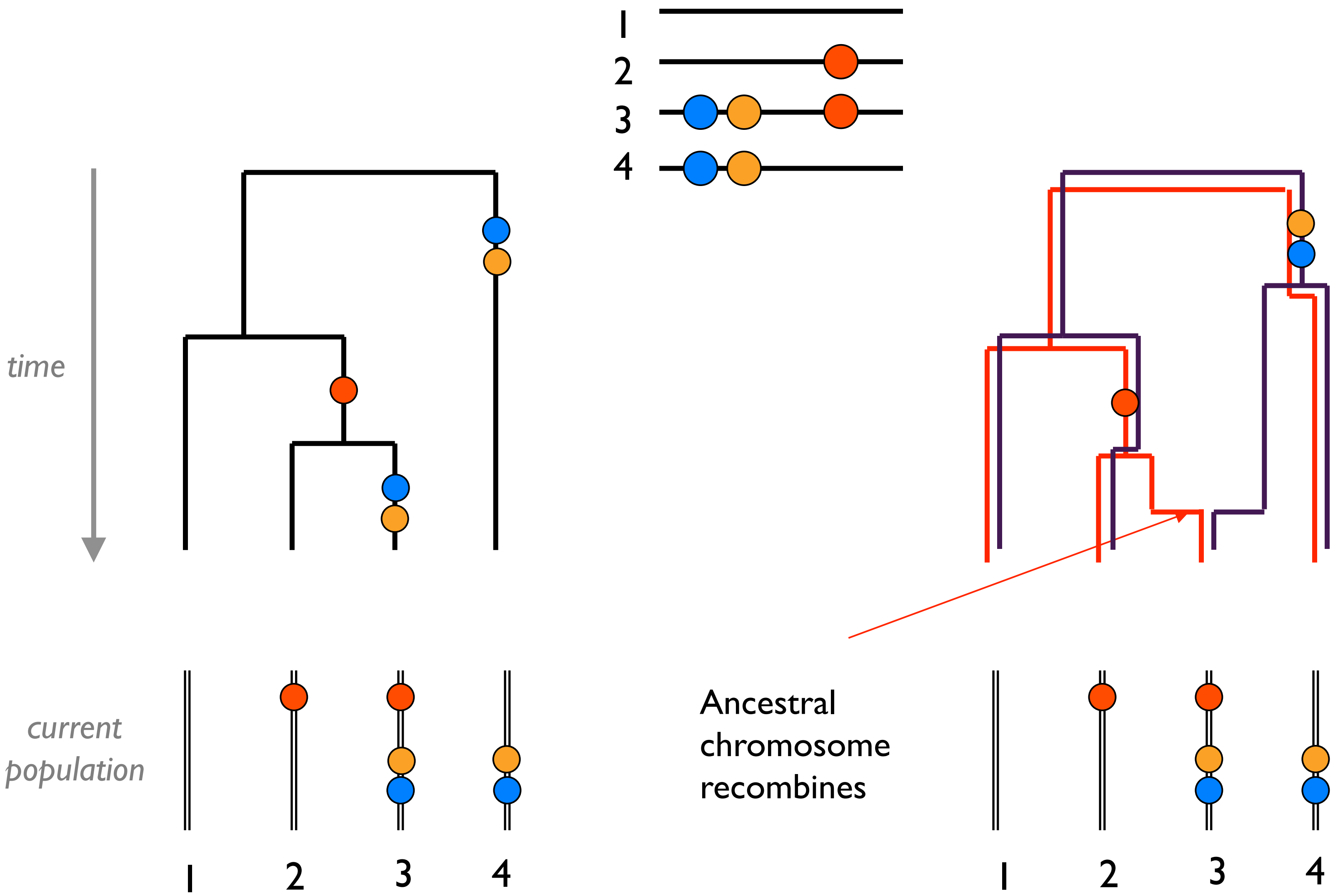
Monomorphic
Obligate intracellular pathogen
Low level of genetic variation
Very low rate of recombination

Adapted from Amine Namouchi, 2012

Outline

- Concepts
  - Recombination and evolution
- Methods currently used to detect recombinant regions
  - Principle and challenges
  - Approaches
- Case study:
  - Superbugs and recombination

Detecting recombination events

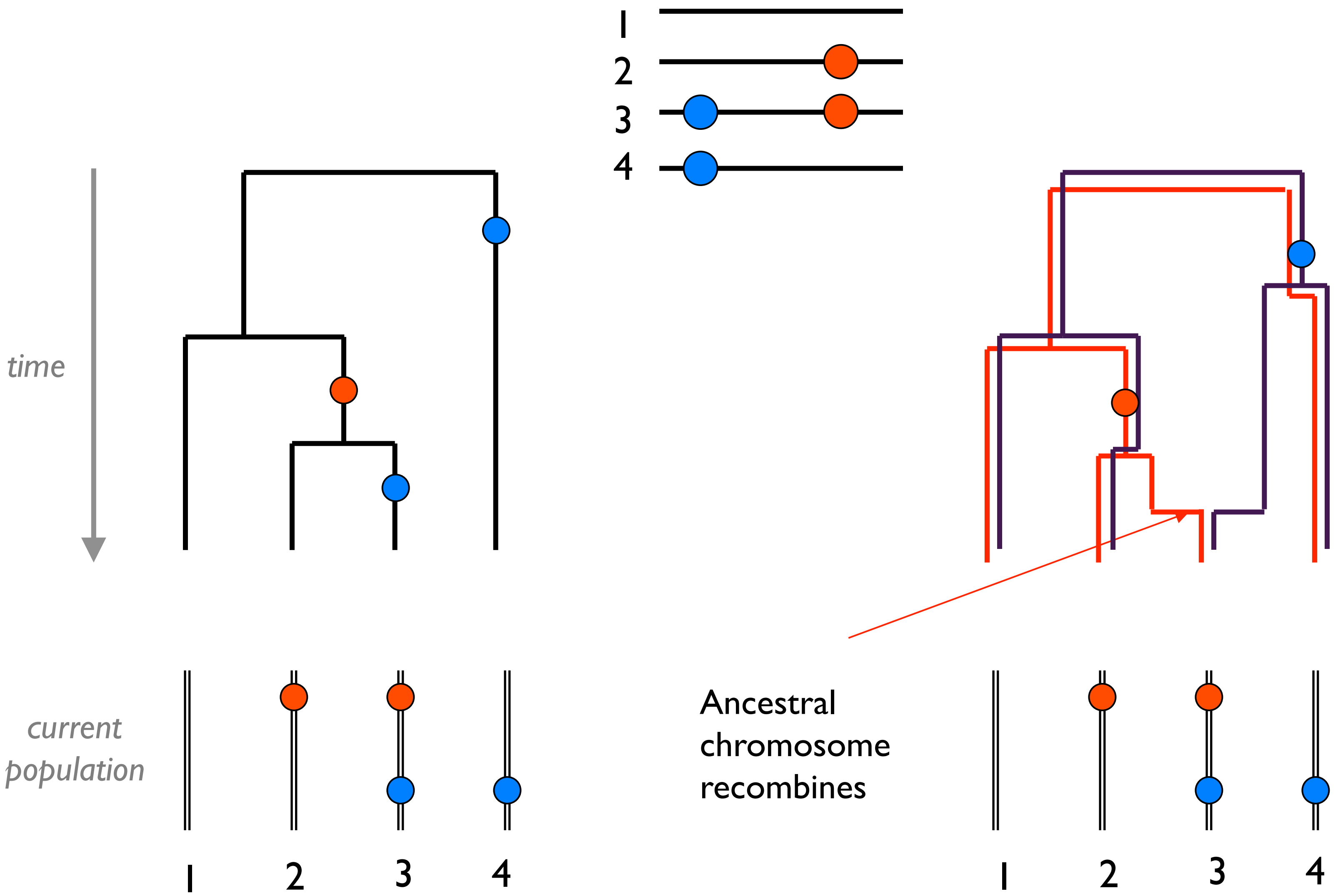


Recurrent mutation

Recombination

Adapted from Simon Myers, 2008

1st challenge: high sequence similarity



Ancestral chromosome recombines

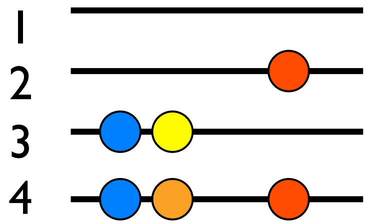
Recurrent mutation

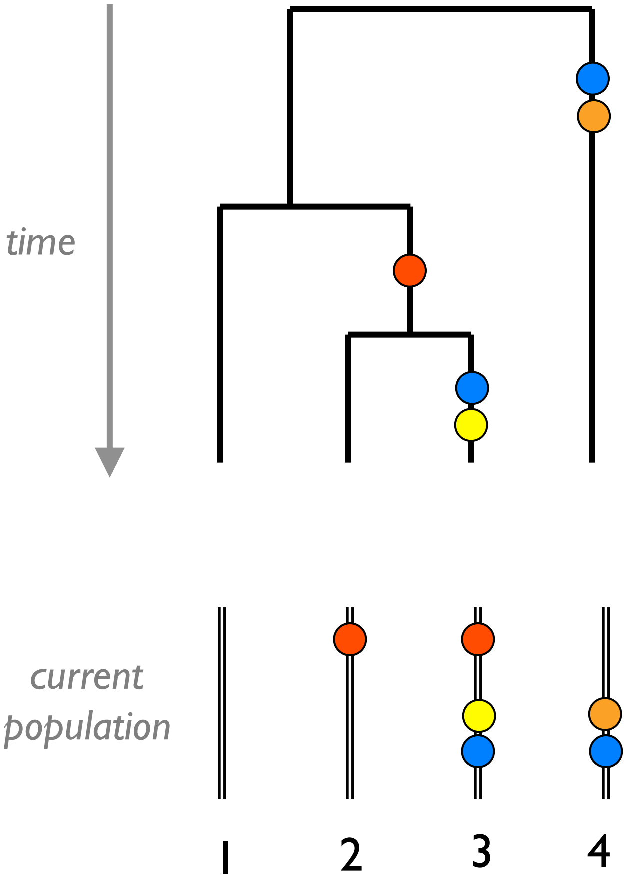
?

Recombination

Adapted from Simon Myers, 2008

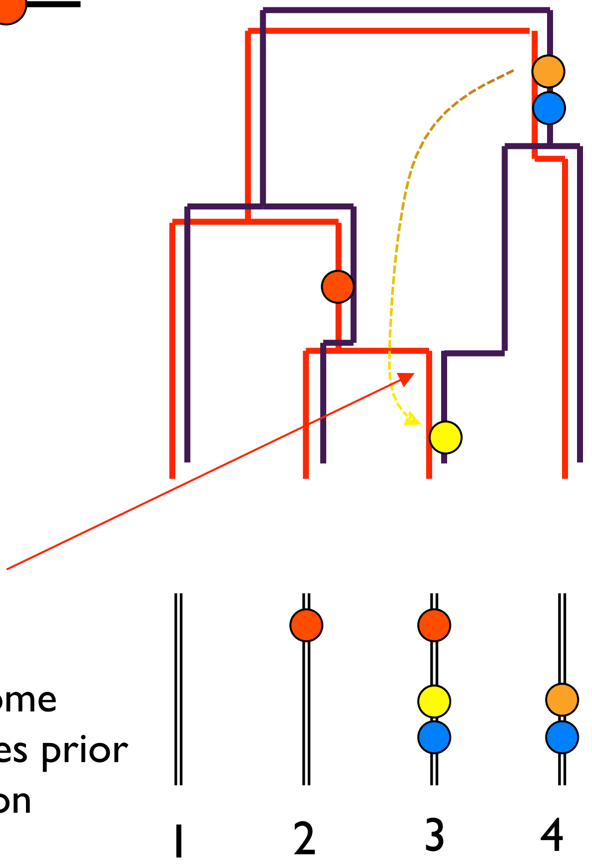
2nd challenge: older recombined regions may have evolved





Recurrent mutation

?

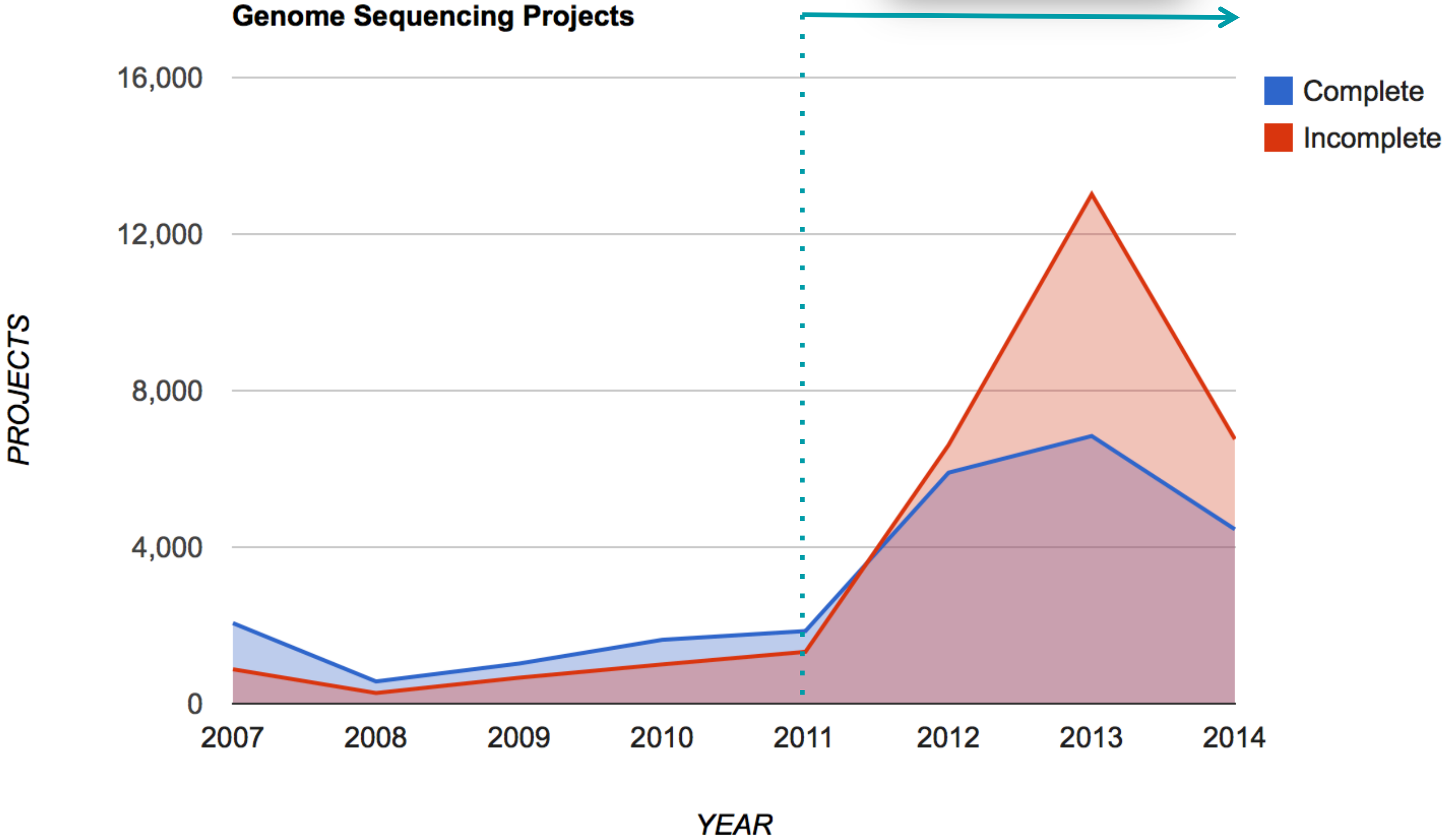


Ancestral chromosome recombines prior to mutation

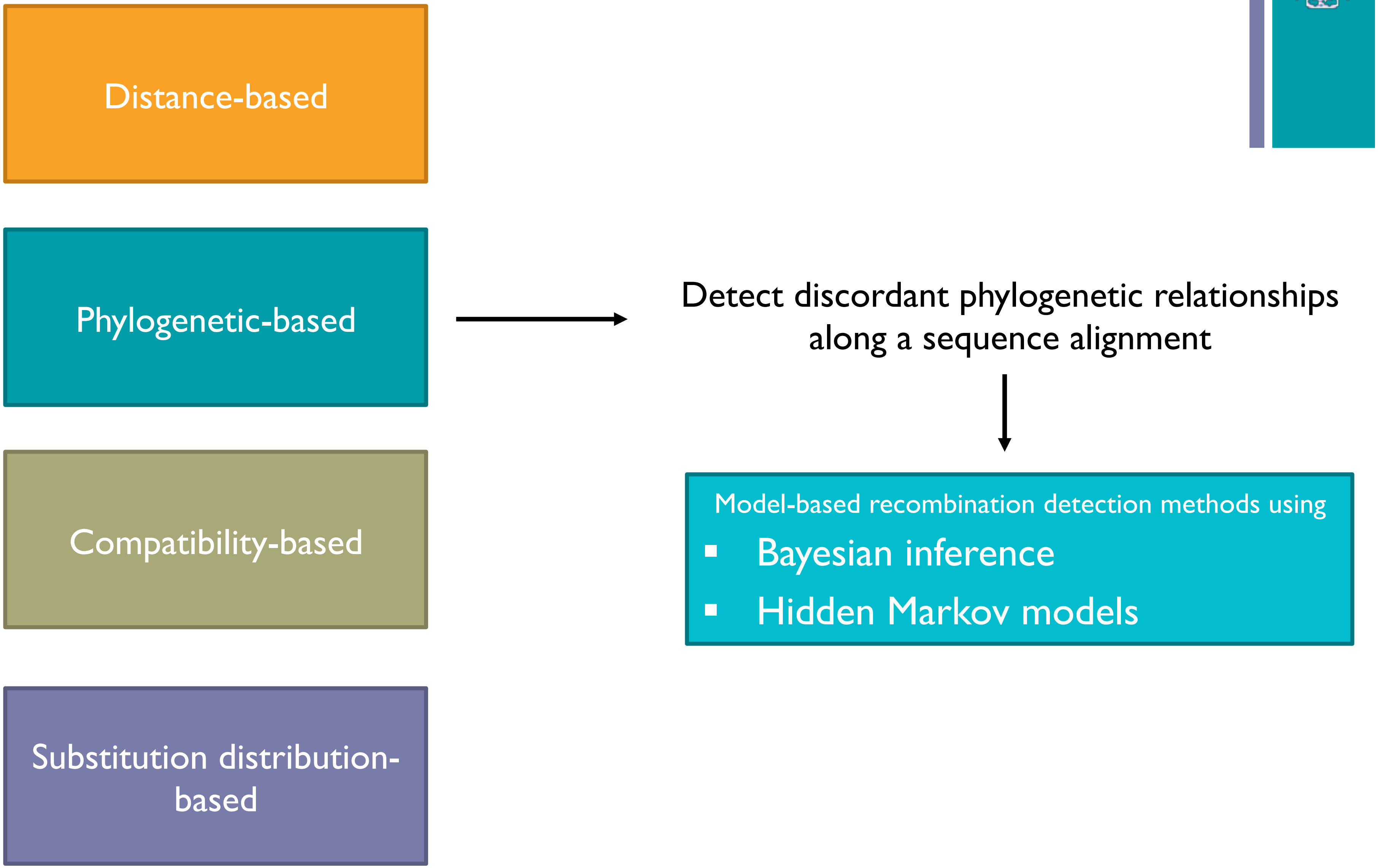
Recombination (older event) + Mutation

Adapted from Simon Myers, 2008

3rd challenge: being in the big data era



Categories of recombination detection programs



Distance-based

Phylogenetic-based

Compatibility-based

Substitution distribution-based

Detect discordant phylogenetic relationships along a sequence alignment

Model-based recombination detection methods using

- Bayesian inference
- Hidden Markov models

Chan CX, Beiko RG, Ragan MA. Detecting recombination in evolving nucleotide sequences. BMC Bioinformatics. 2006

ClonalFrame

+

- estimates the clonal relationships between the members of a dataset
- estimates the chromosomal position of homologous recombination events that have disrupted the clonal inheritance
- used to be a gold standard method

-

- restricted to small datasets (theoretically scales up to whole genomes close to a 100)
- does not model the origin of genetic imports
- not appropriate if the recombination rate is too high

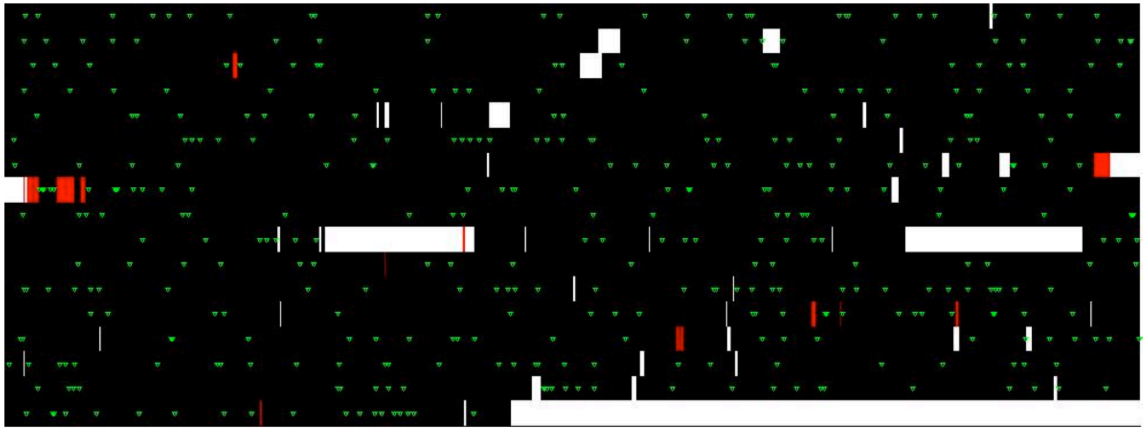


Figure 6: Example of genomic representation of the mutation and recombination events for a whole genome alignment

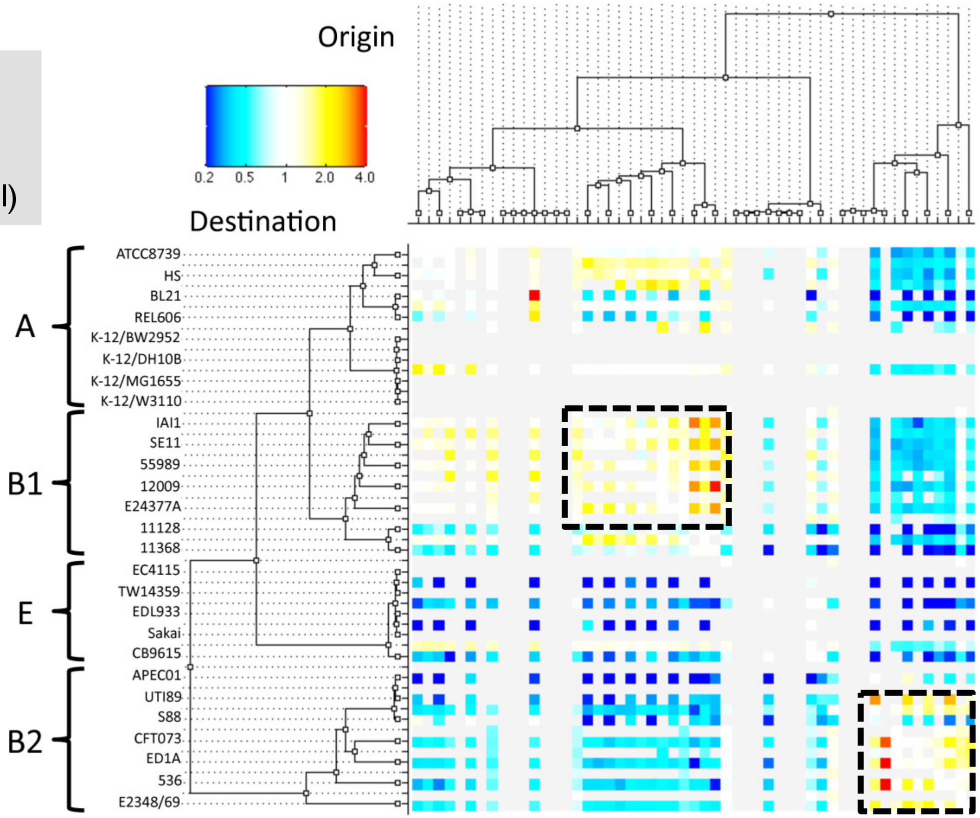
Didelot X, Falush D. Inference of bacterial microevolution using multilocus sequence data. Genetics. 2007;175:1251–1266.

ClonalOrigin: identify flux of recombination

- models bacterial recombination as an event from a specific donor to a specific recipient

- but shares most of the drawbacks of ClonalFrame
- also requires a clonal genealogy (provided by ClonalFrame)

Number of recombination events inferred by ClonalOrigin relative to its expectation under the inferred recombination rate (predicted by model)



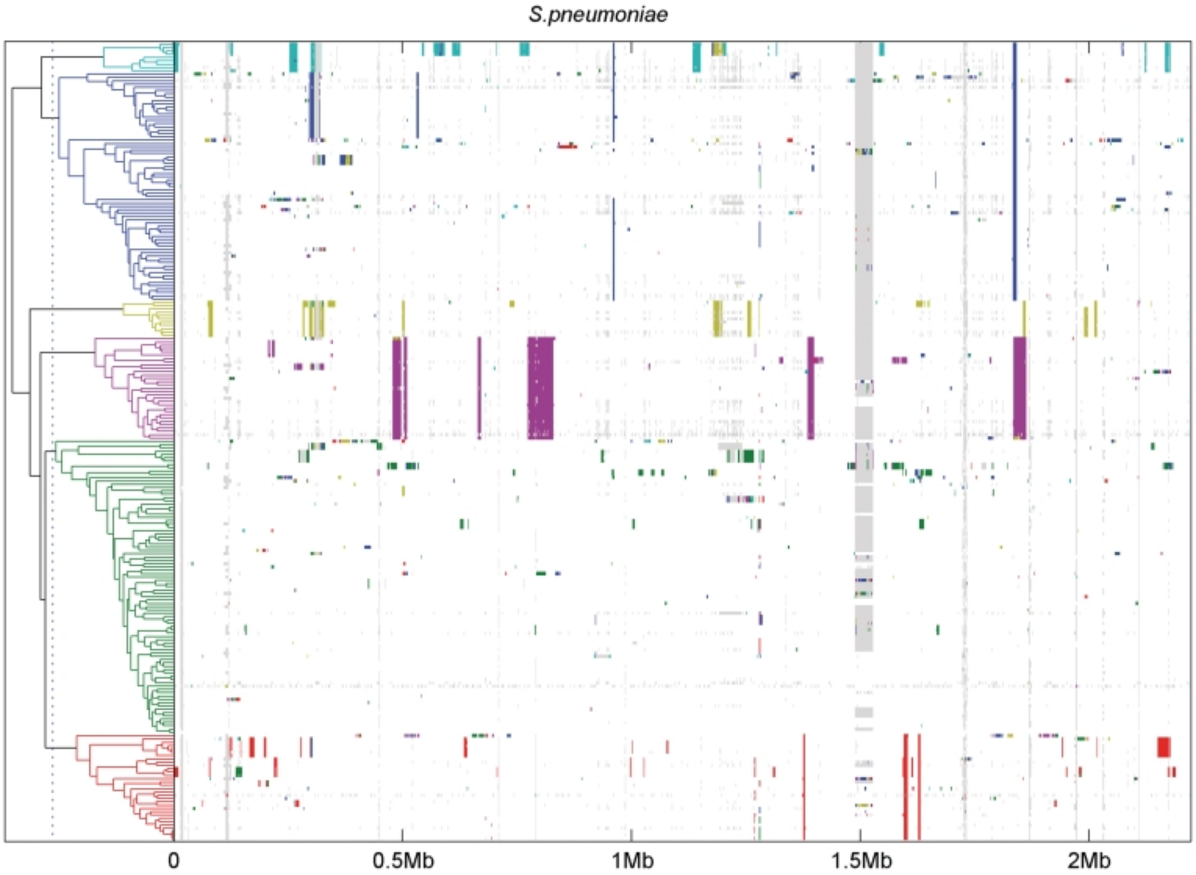
Didelot X, Lawson D, Darling A, Falush D. Inference of homologous recombination in bacteria using whole-genome sequences. Genetics. 2010;186:1435–1449

Didelot X, Méric G, Falush D, Darling A. Impact of homologous and non-homologous recombination in the genomic evolution of Escherichia coli. BMC Genomics 2012;13:256

BRATNextGen

- based on Bayesian change-point clustering model + permutation re-sampling procedure
- adapted to very large datasets (200-300 genomes and more)

- cannot model direction of recombination events (relative to cluster size)
- expects a more or less clonal population



Color = Major cluster in which recombination is observed

Marttinen P, et al. Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res. 2011.

Chromosome painting (Chromo Painter and FineSTRUCTURE)

+

- each genome is reconstructed using DNA chunks donated by other genomes
- well suited for highly recombinogenic (non-clonal) datasets (*i.e. H. pylori*)
- can model recombination events between the observed sequences

-

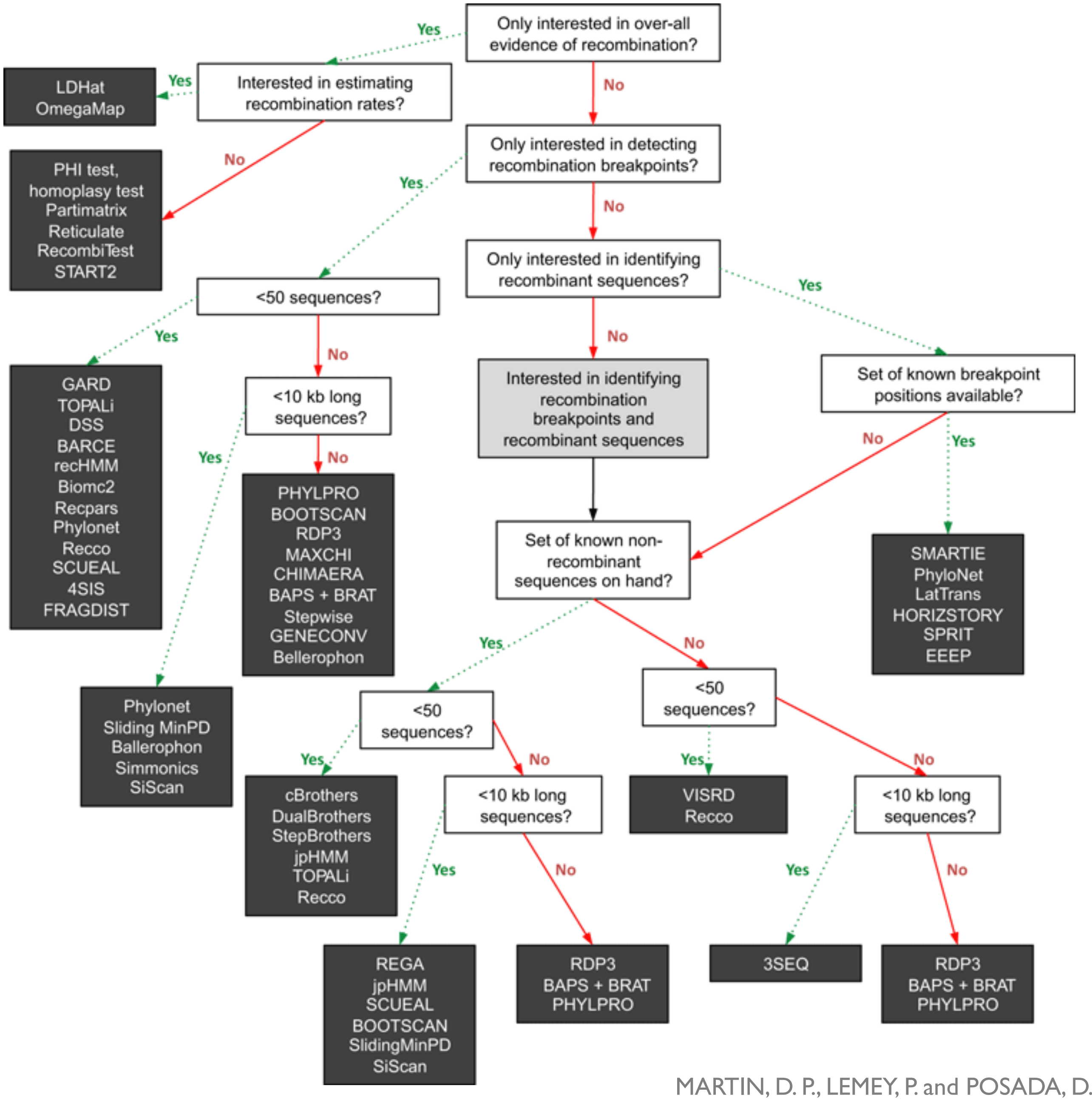
- initially developed for eukaryotes
- possible uncertainty when multiple haplotypes are equally close

SUBGROUPS

STRAINS

Yahara K. *et al.* Chromosome Painting In Silico in a Bacterial Species Reveals Fine Population Structure. Mol Biol Evol (2013)

Choosing the appropriate recombination analysis program



Only interested in over-all evidence of recombination?

Interested in estimating recombination rates?

LDHat
OmegaMap

PHI test,
homoplasy test
Partimatrix
Reticulate
RecombiTest
START2

Only interested in detecting recombination breakpoints?

<50 sequences?

GARD
TOPALi
DSS
BARCE
recHMM
Biomc2
Recpars
Phylonet
Recco
SCUEAL
4SIS
FRAGDIST

<10 kb long sequences?

Phylonet
Sliding MinPD
Ballerophon
Simmonics
SiScan

PHYLPRO
BOOTSCAN
RDP3
MAXCHI
CHIMAERA
BAPS + BRAT
Stepwise
GENECONV
Bellerophon

Only interested in identifying recombinant sequences?

Set of known breakpoint positions available?

SMARTIE
PhyloNet
LatTrans
HORIZSTORY
SPRIT
EEEP

Interested in identifying recombination breakpoints and recombinant sequences

Set of known non-recombinant sequences on hand?

<50 sequences?

cBrothers
DualBrothers
StepBrothers
jpHMM
TOPALi
Recco

<10 kb long sequences?

REGA
jpHMM
SCUEAL
BOOTSCAN
SlidingMinPD
SiScan

RDP3
BAPS + BRAT
PHYLPRO

<50 sequences?

VISRD
Recco

<10 kb long sequences?

3SEQ

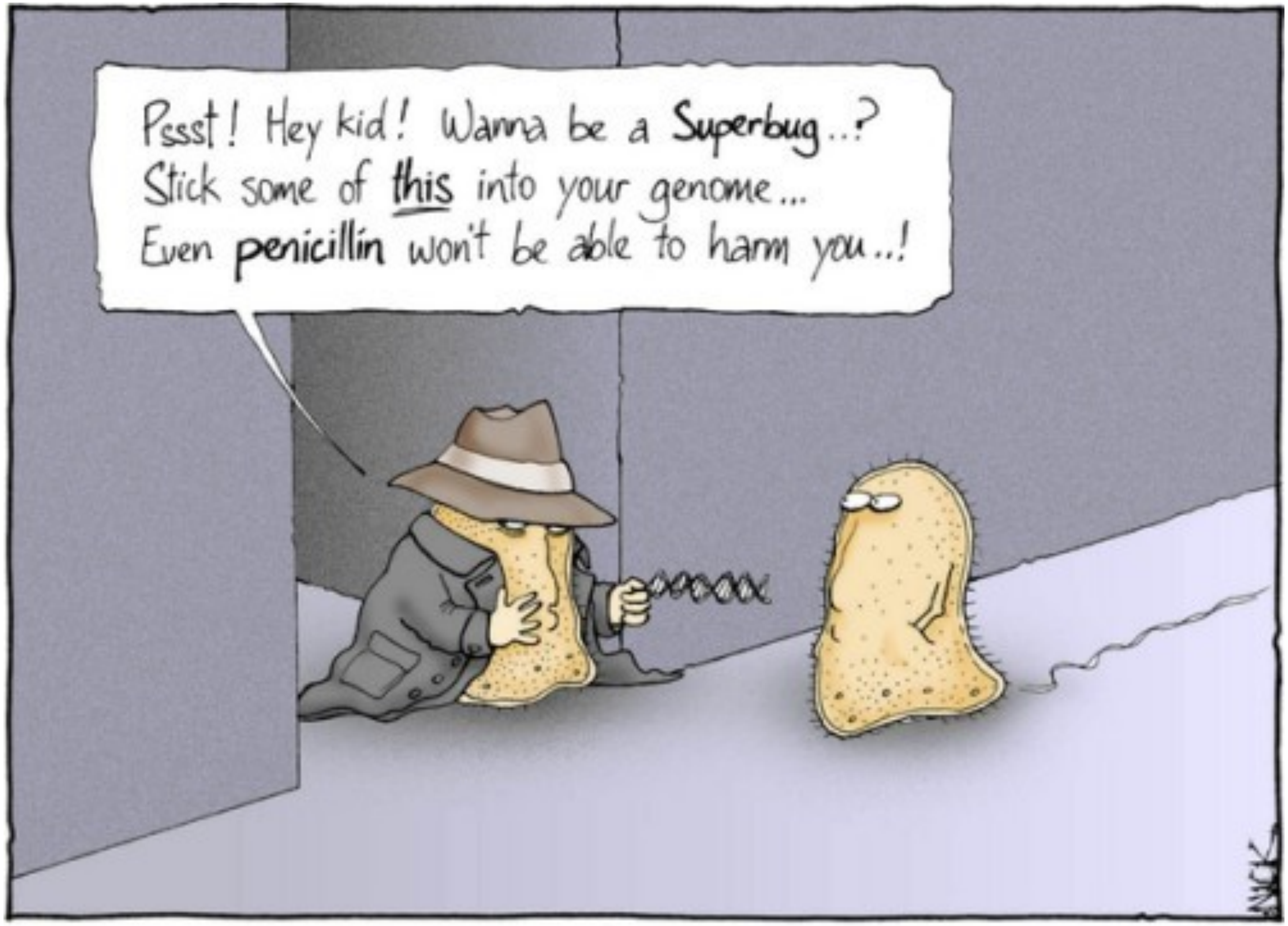
RDP3
BAPS + BRAT
PHYLPRO

MARTIN, D. P., LEMEY, P. and POSADA, D. (2011), Analysing recombination in nucleotide sequences. Molecular Ecology Resources, 11: 943–955

Outline

- Concepts
  - Recombination and evolution
- Methods currently used to detect recombinant regions
  - Principal and challenges
  - Approaches
- Case study:
  - Superbugs and recombination

Superbugs: rise of the antibiotic resistance



It was on a short-cut through the hospital kitchens that Albert was first approached by a member of the Antibiotic Resistance.

Superbugs…

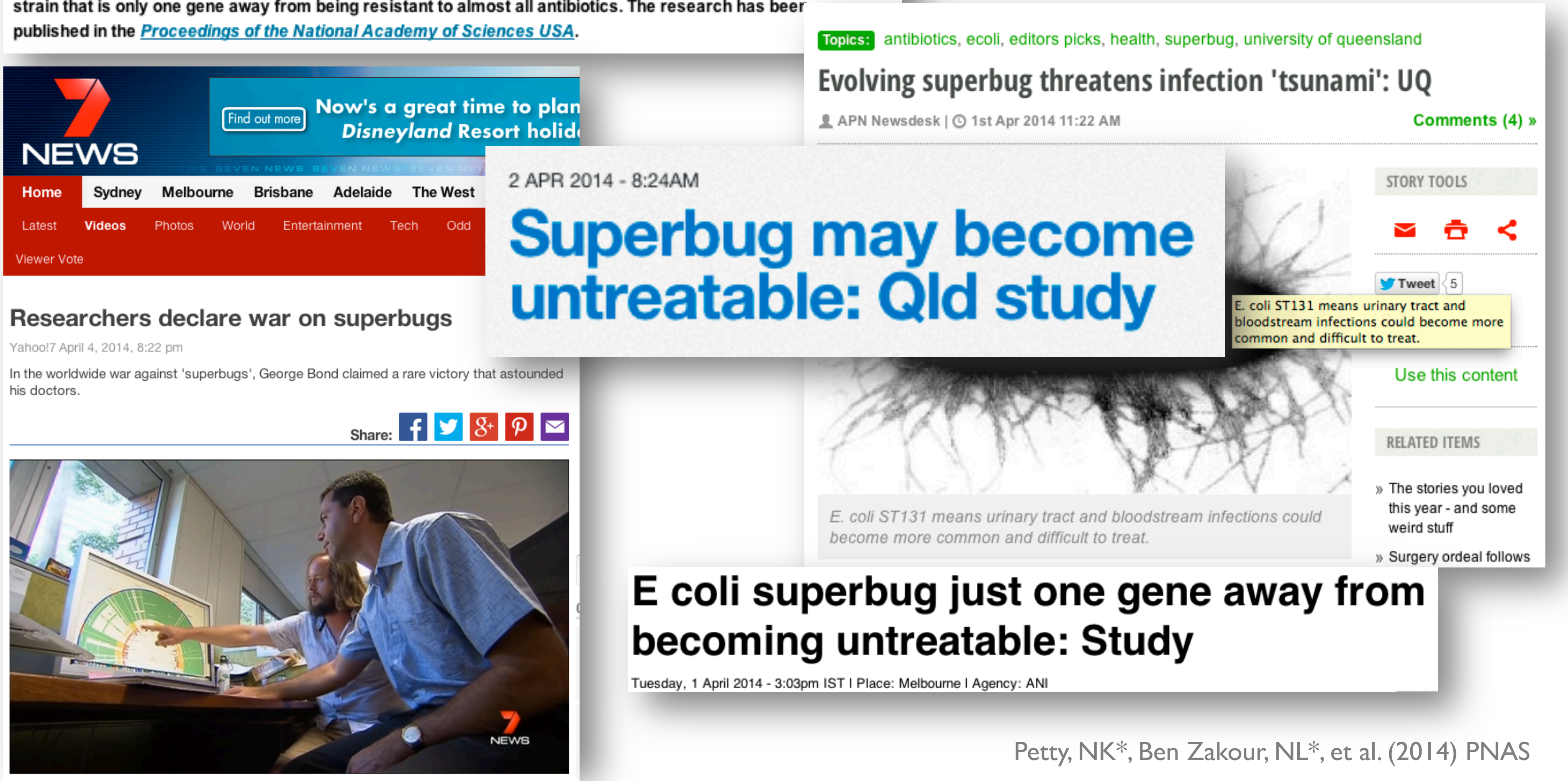
Home > News > *E. coli* strain close to complete resistance

Posted: Apr 2, 2014

Topics: Life science & clinical diagnostics instruments > Microbiology

E. coli strain close to complete resistance

An international study, led by The University of Queensland (UQ), has been tracking a multidrug-resistant *E. coli* strain that is only one gene away from being resistant to almost all antibiotics. The research has been published in the *Proceedings of the National Academy of Sciences USA*.



Researchers declare war on superbugs

Yahoo!7 April 4, 2014, 8:22 pm

In the worldwide war against 'superbugs', George Bond claimed a rare victory that astounded his doctors.

2 APR 2014 - 8:24AM

Superbug may become untreatable: Qld study

Topics: antibiotics, ecoli, editors picks, health, superbug, university of queensland

Evolving superbug threatens infection 'tsunami': UQ

APN Newsdesk | 1st Apr 2014 11:22 AM

E. coli ST131 means urinary tract and bloodstream infections could become more common and difficult to treat.

» The stories you loved this year - and some weird stuff

» Surgery ordeal follows

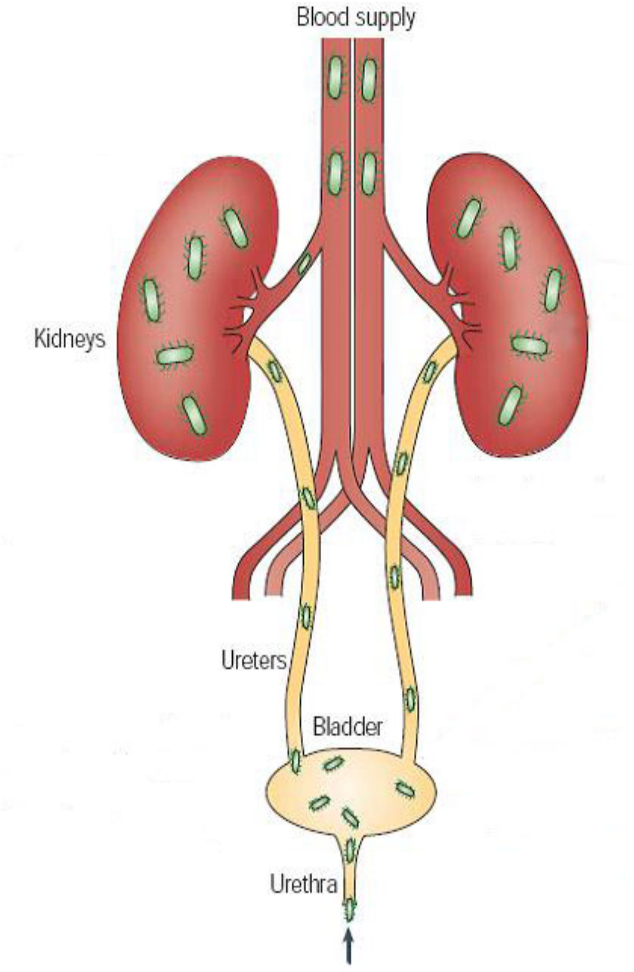
E coli superbug just one gene away from becoming untreatable: Study

Tuesday, 1 April 2014 - 3:03pm IST | Place: Melbourne | Agency: ANI

E. coli ST131 means urinary tract and bloodstream infections could become more common and difficult to treat.

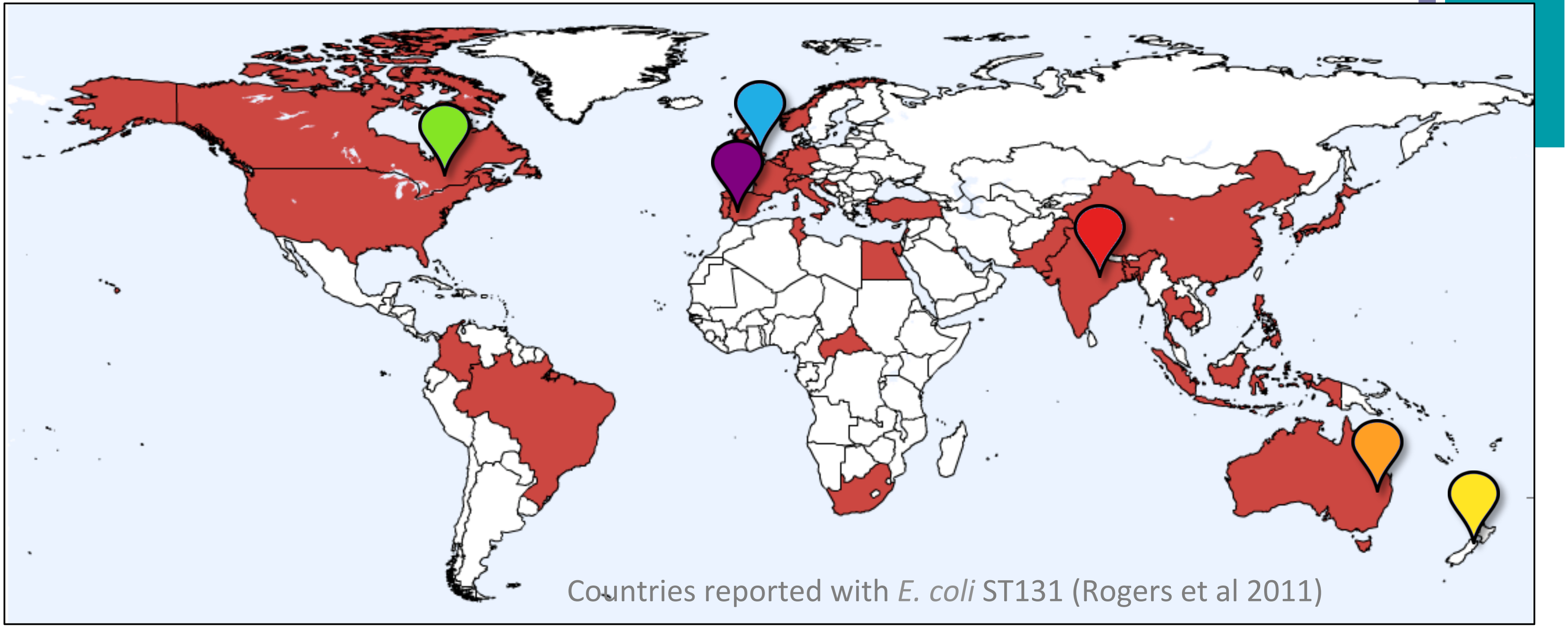
Petty, NK*, Ben Zakour, NL*, et al. (2014) PNAS

… And Urinary Tract Infections



- UTI is very common, especially among women and the elderly
- Multi-drug resistant organisms more common (XDR bacteria)
- Recurrent infections, difficult to treat

E. coli ST131: a multidrug resistant clone gone global



Countries reported with *E. coli* ST131 (Rogers et al 2011)

Escherichia coli **ST131**

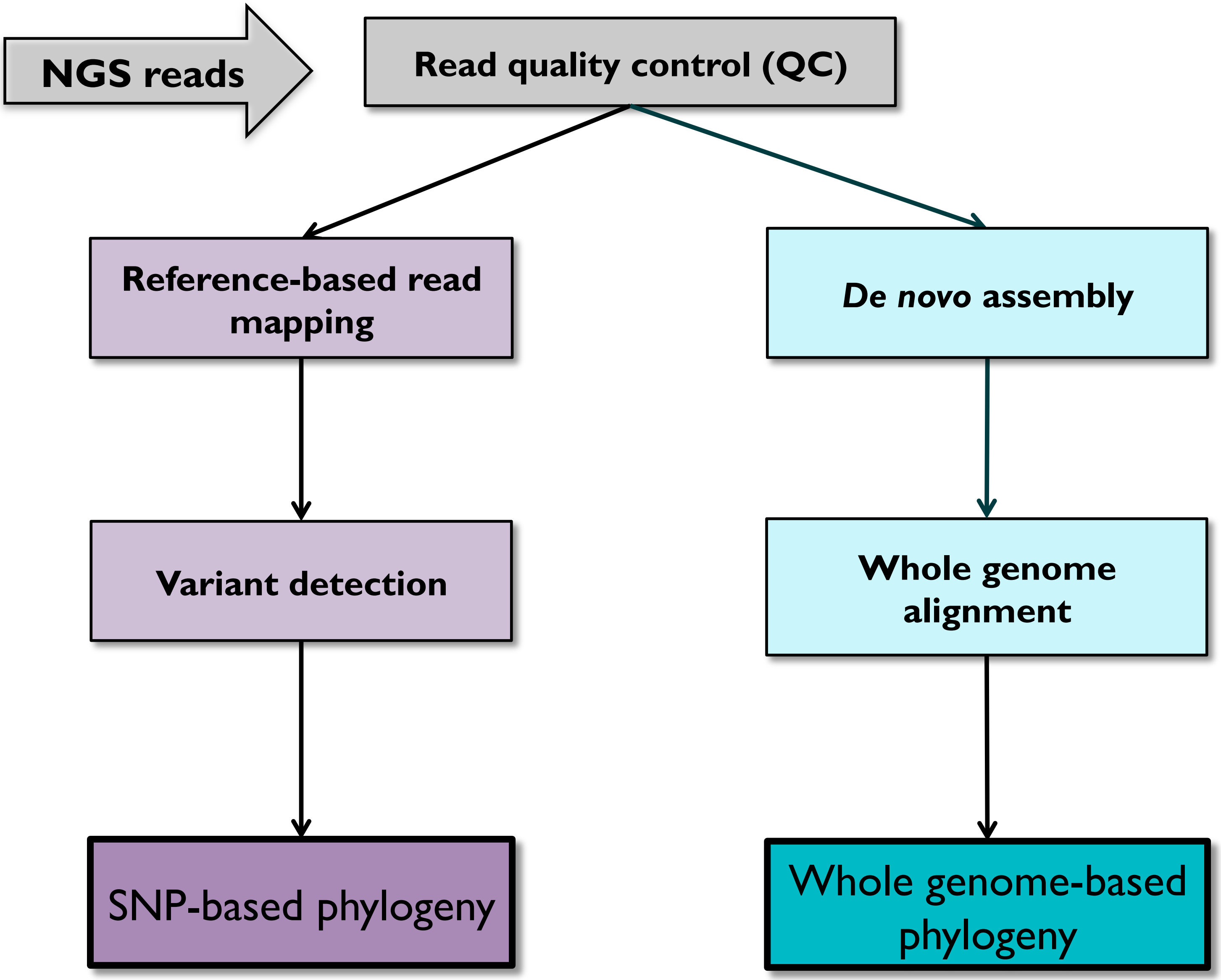
- globally spread since 2008
- urinary tract and bloodstream infections
- multidrug resistant (ESBL CTX-M-15[a])
- high prevalence of virulence factors

95 ST131 uropathogenic *E. coli* (UPEC) strains, spanning 2000-2011

[a] CTX-M: active on CefoTaXime, first isolated in Munich

Petty, NK*, Ben Zakour, NL*, et al. (2014) PNAS

Banzai genomic pipeline (M. Stanton-Cook, E. Skippington)



NGS reads

Read quality control (QC)

Reference-based read mapping

Variant detection

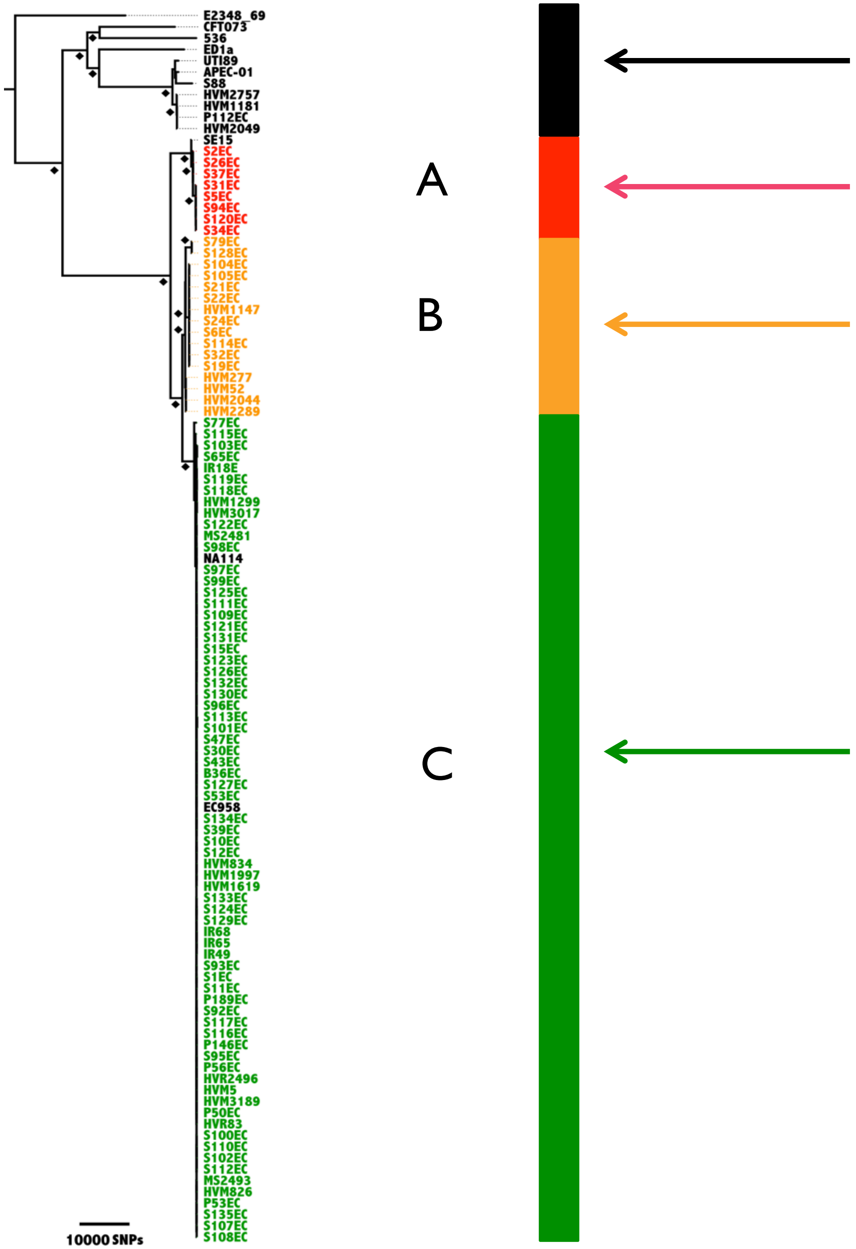
SNP-based phylogeny

***De novo* assembly**

Whole genome alignment

Whole genome-based phylogeny

Global phylogeny of the ST131 collection



non ST131 UPEC strains

most divergent clade (~7,000 and ~8,900 SNPs from B and C, respectively)

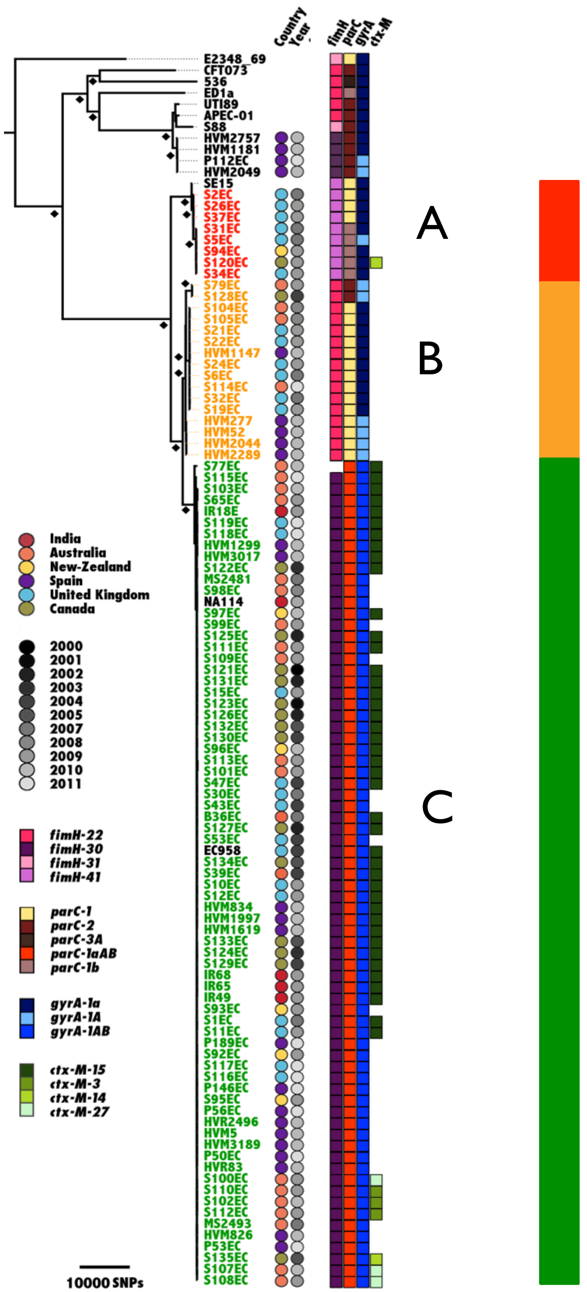
clade B is very similar to clade C (~2,900 SNPs apart)

clade C is the dominant clade (70% of strains sequenced in this study)

Maximum Likelihood Tree
Build using 142,950 SNPs
95 ST131 strains +
15 representative non-ST131
UPEC genomes

Petty, NK*, Ben Zakour, NL*, et al. (2014) PNAS

Lack of temporal and geographical clustering in ST131



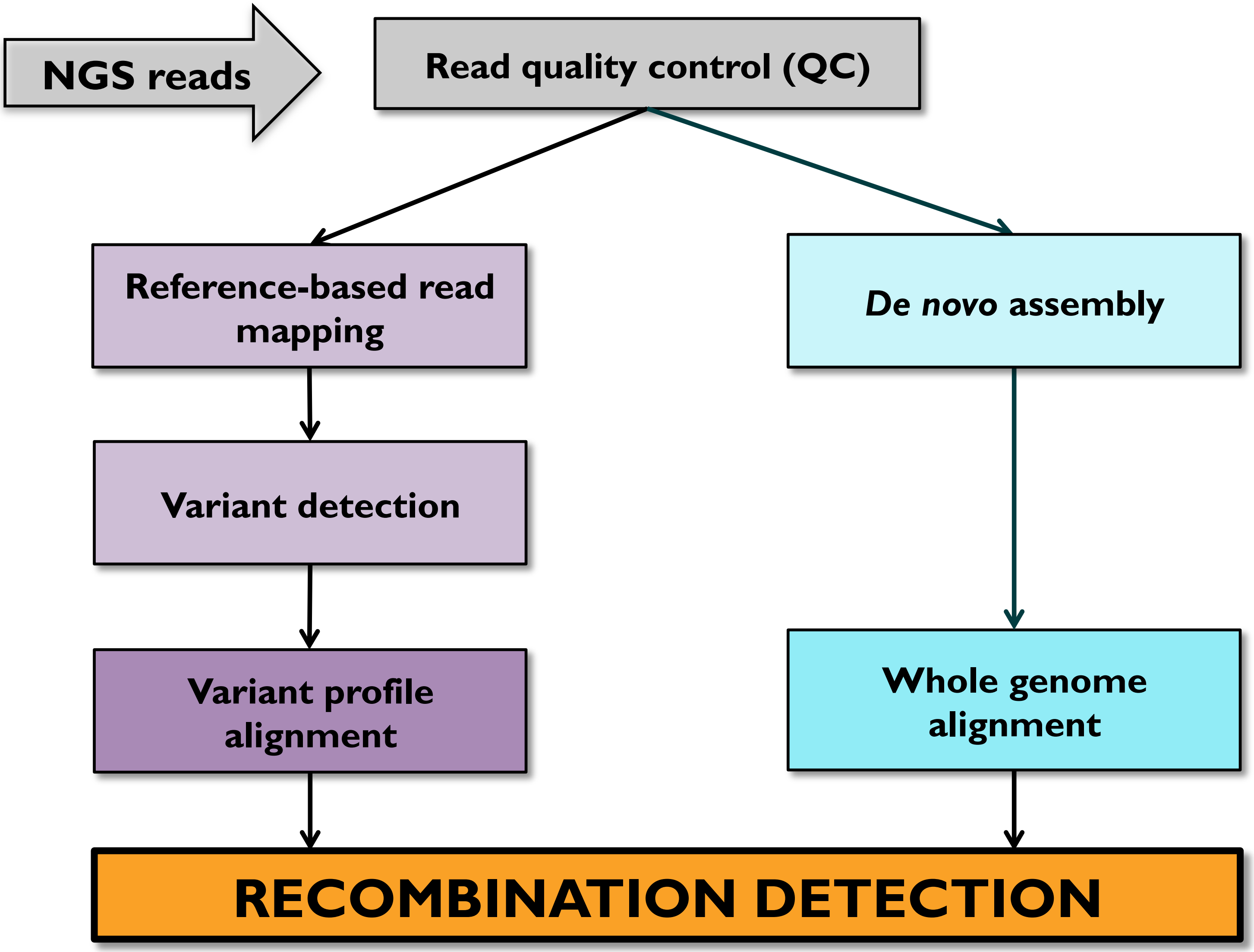
Little temporal or geographical association between clades

Clade specific variants of virulence/resistance determinants: *fimH, parC, gyrA, ctxM…*

Maximum Likelihood Tree
Build using 142,950 SNPs
95 ST131 strains +
15 representative non-ST131
UPEC genomes

Petty, NK*, Ben Zakour, NL*, et al. (2014) PNAS

Possible inputs for recombination detection (in Banzai)



NGS reads

Read quality control (QC)

Reference-based read mapping

Variant detection

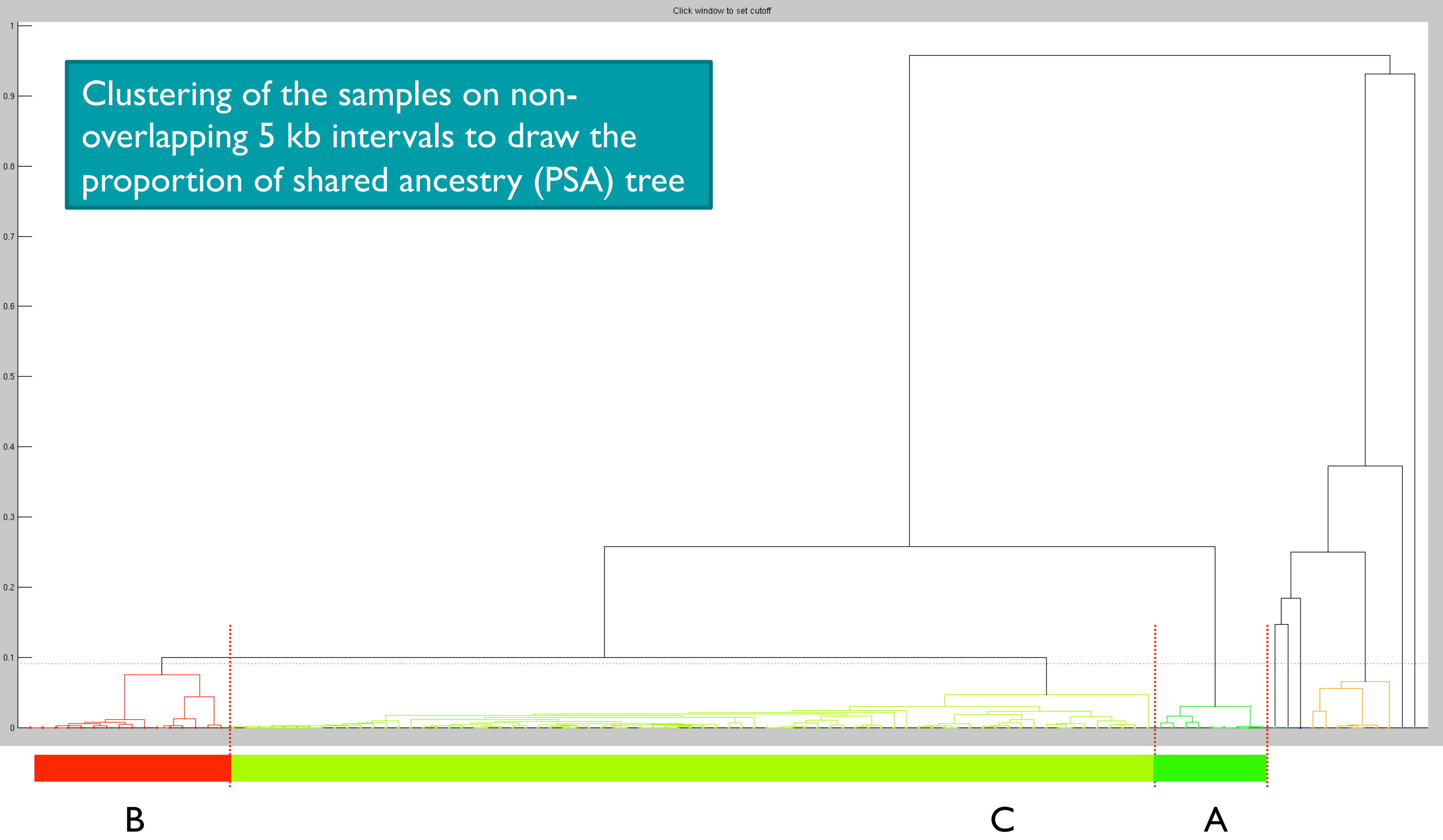
Variant profile alignment

***De novo* assembly**

Whole genome alignment

RECOMBINATION DETECTION

BRATNextGen: Proportion of Shared Ancestry hierarchical tree



Clustering of the samples on non-overlapping 5 kb intervals to draw the proportion of shared ancestry (PSA) tree

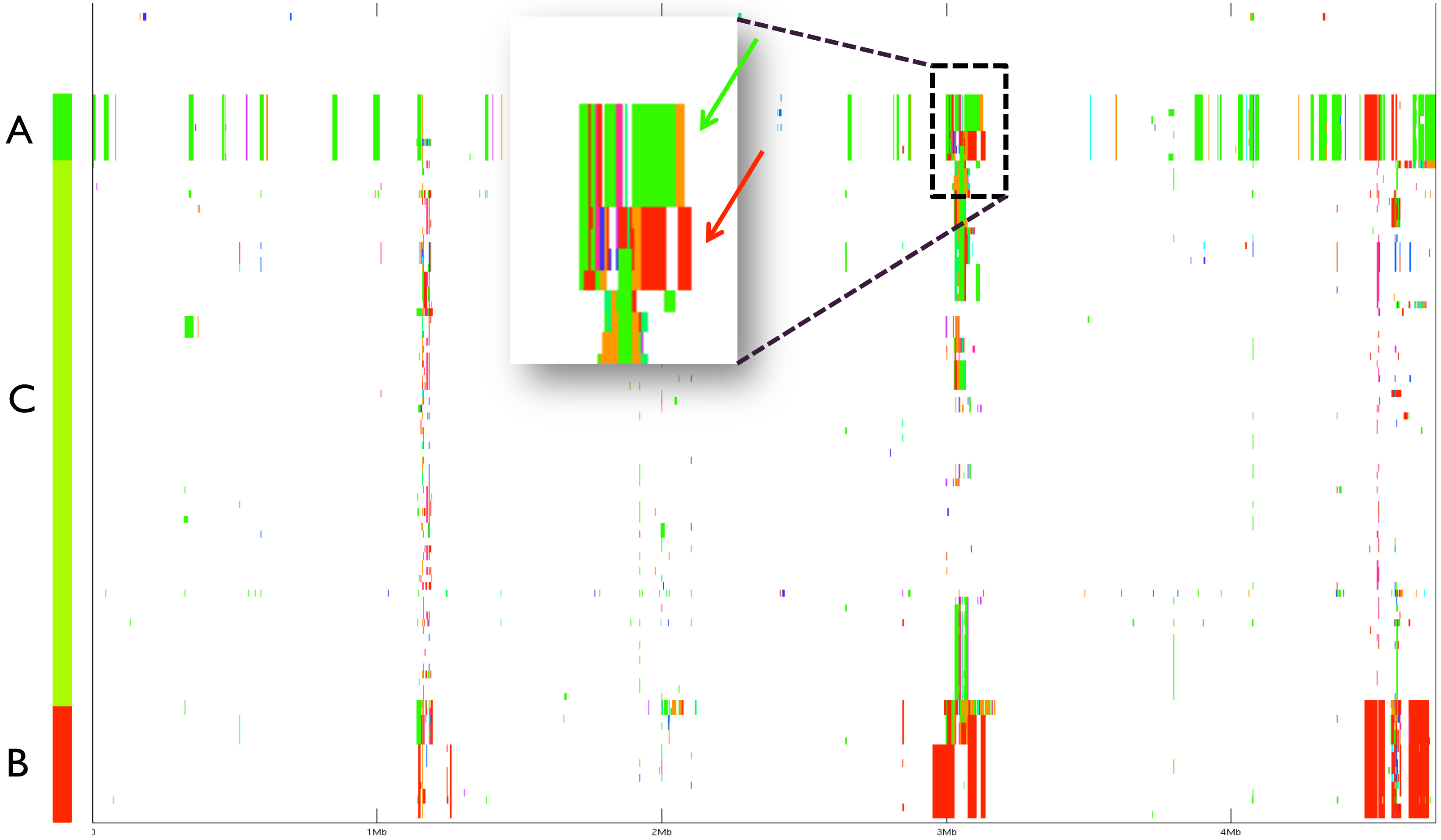
BRATNextGen: recombination map

BratNextGen predicted 137 combined regions eq. 0.94 Mb (1/3 of the core genome)

Example of clade-specific blocks of recombination

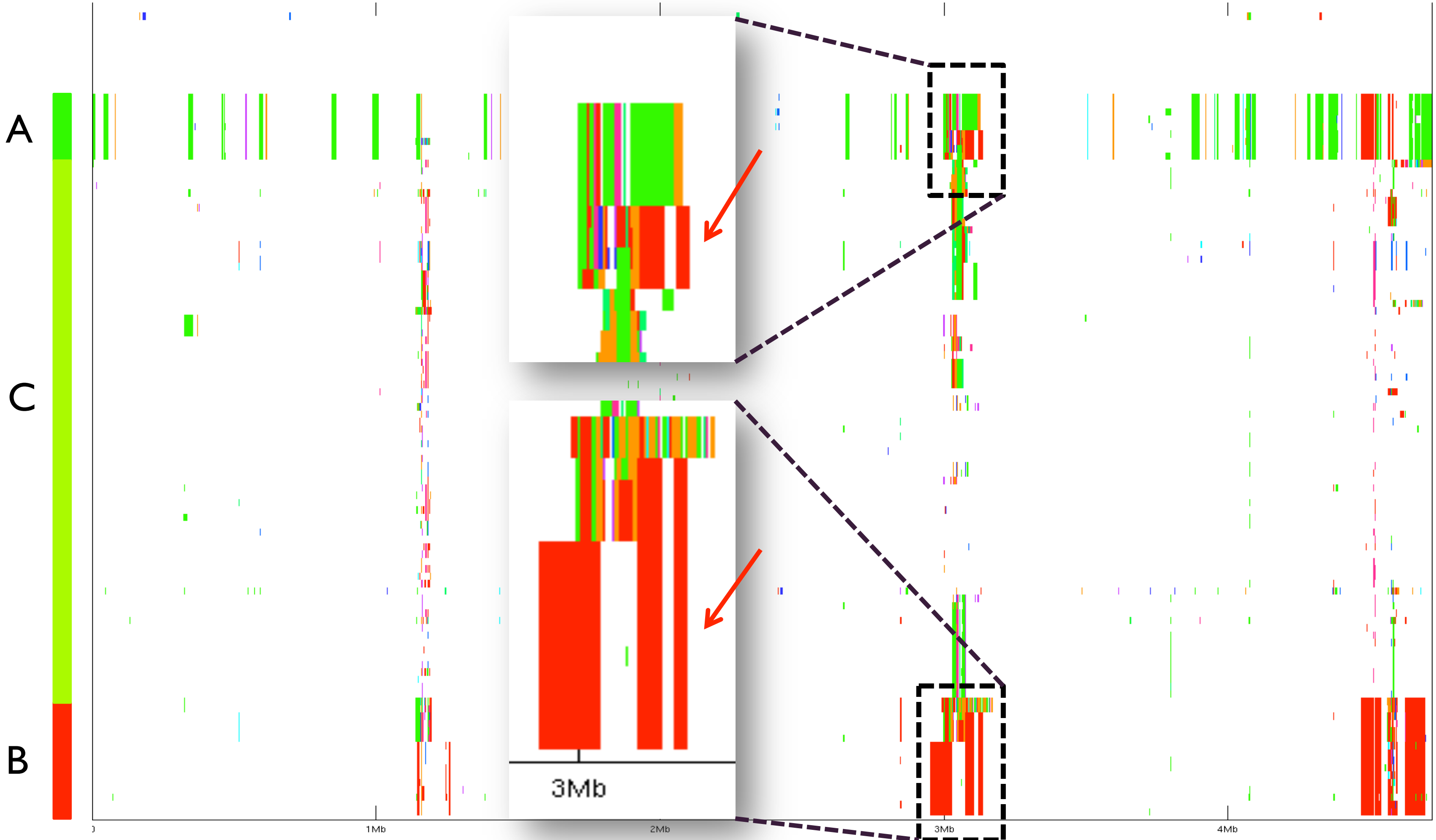
Estimating the amount of recombination

Example of distinct recombination events within-clade



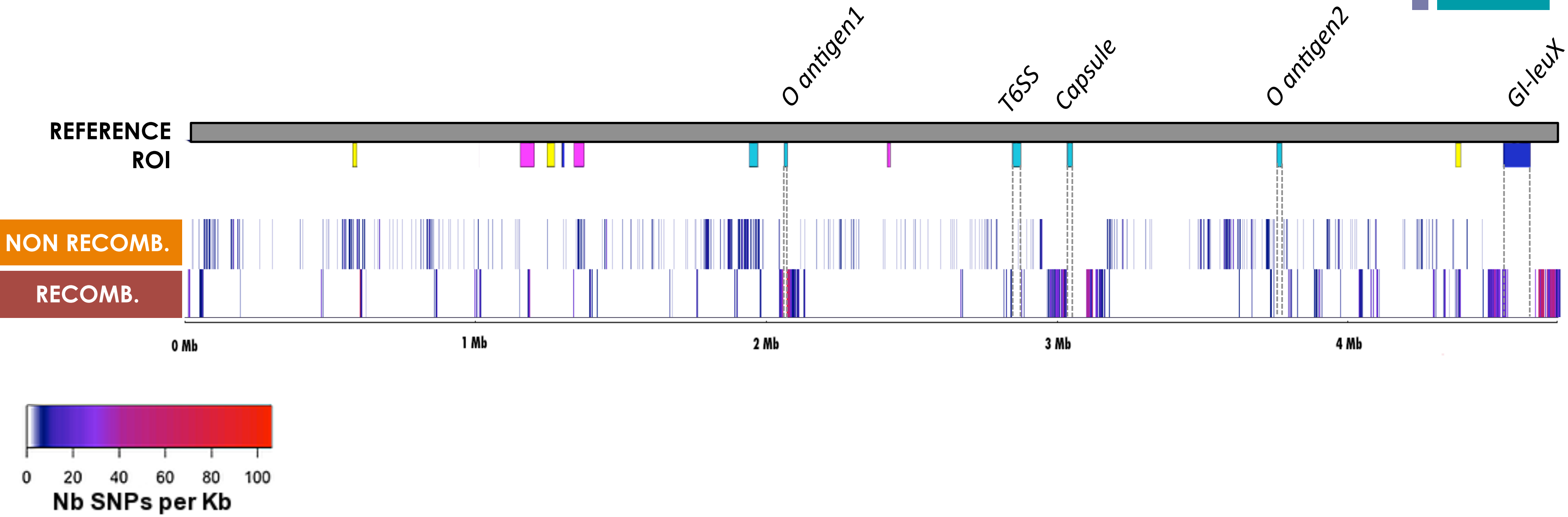
Estimating the amount of recombination

Example of similar recombination events between-clade



Recombination regions distribution is not random

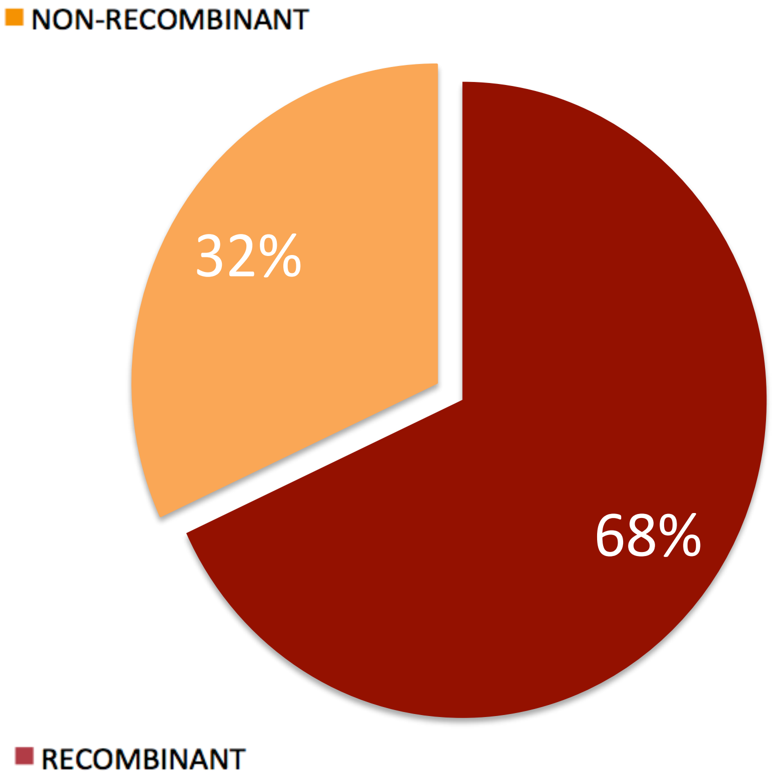
Distribution of ST131-only core SNPs in non-recombinant vs recombinant regions

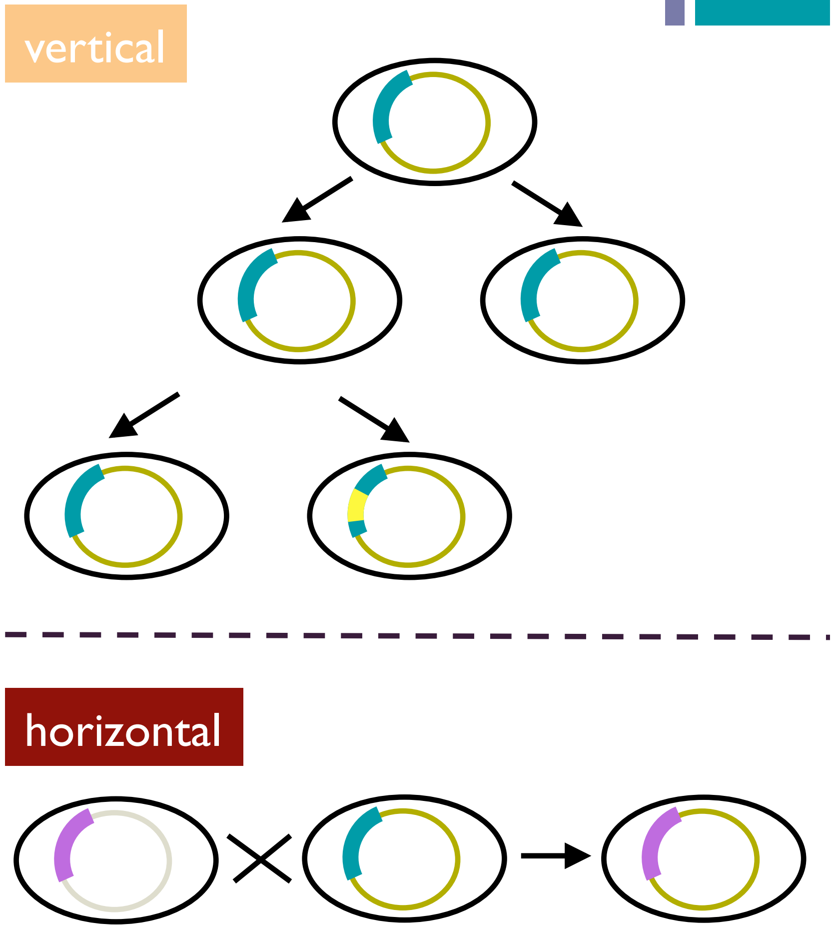


- Statistical analysis suggested association of recombinant regions with regions of interests (mobile genetic elements, genomic islands, *etc.*)
- Also established in other species (Everitt et al. Mobile elements drive recombination hotspots in the core genome of *Staphylococcus aureus*. Nature Comm. 2014)

A large proportion of the SNPs has been introduced through recombination

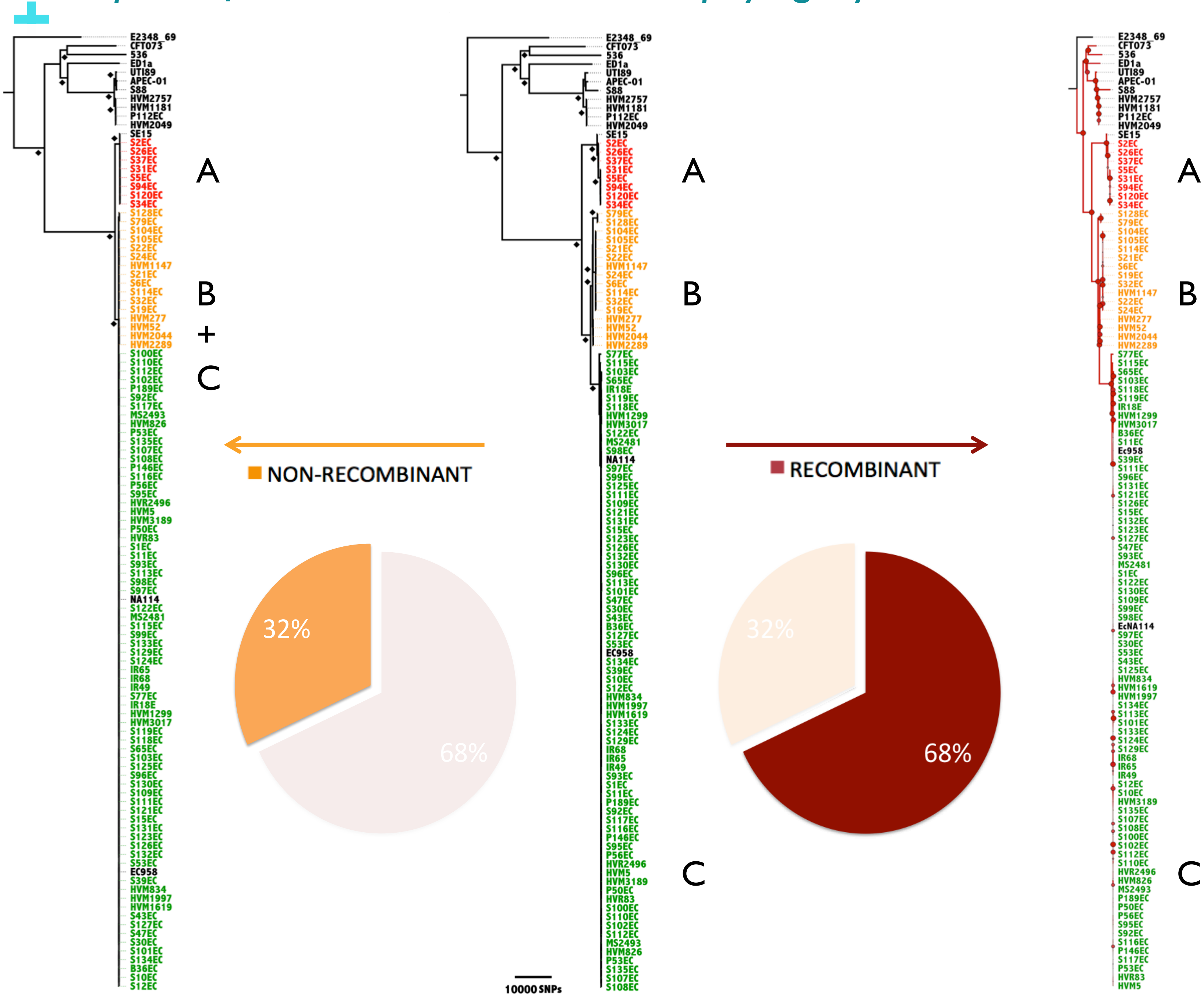
Proportion of ST131-specific SNPs in non-recombinant and recombinant regions





Petty, NK*, Ben Zakour, NL*, et al. (2014) PNAS

Impact of recombination on ST131 phylogeny reconstruction

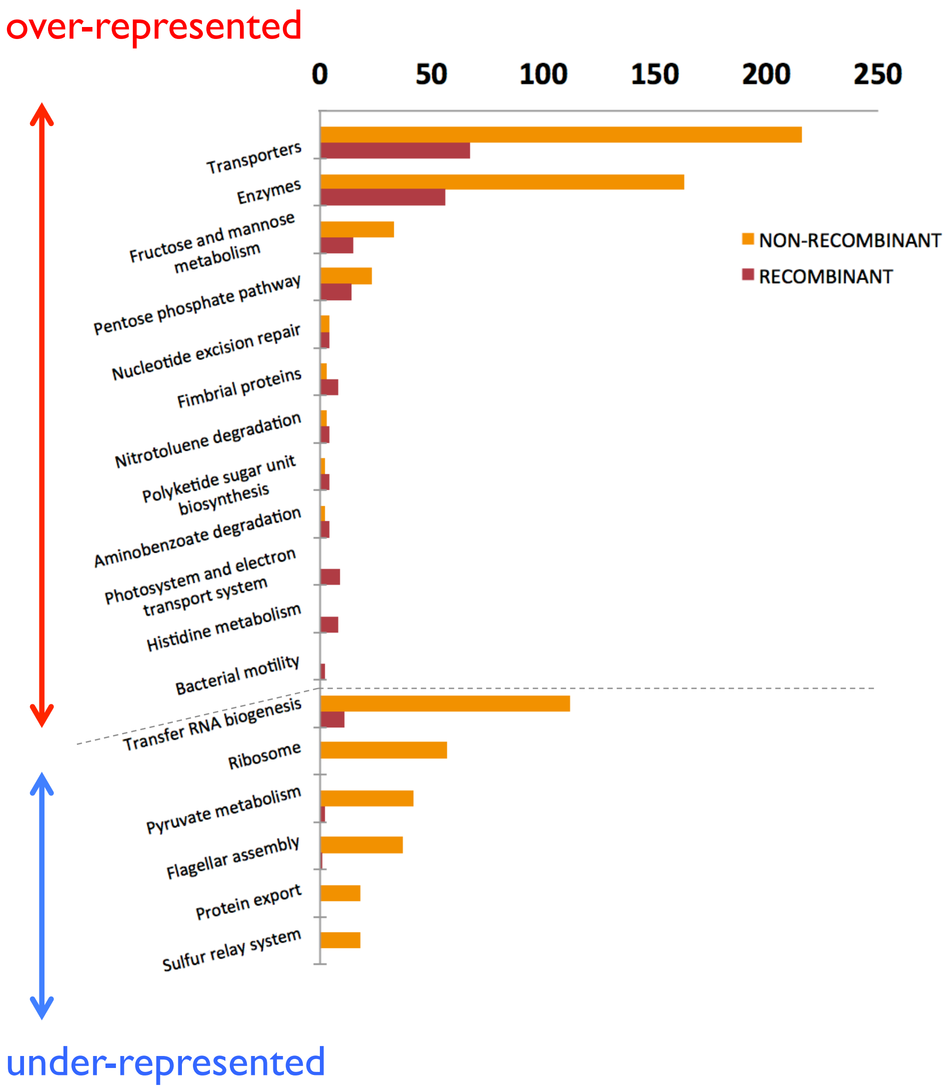


🟧 NON-RECOMBINANT

🟥 RECOMBINANT

Impact of recombination on ST131 functional adaptation

KEGG functional categories significantly different



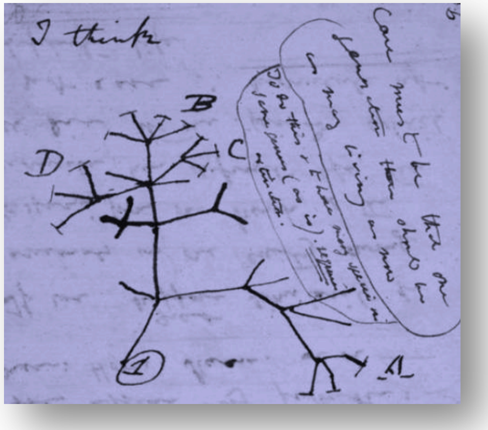
Example of virulence associated functions

- Resistance to FQ (*parC* variant)
- Salt-resistance
- Iron transporter
- Type 2 and 6 Secretion Systems
- Capsule biosynthesis
- Type I fimbriae
- Flagellar locus

> 900 genes affected by recombination

Petty, NK*, Ben Zakour, NL*, et al. (2014) PNAS

Summary



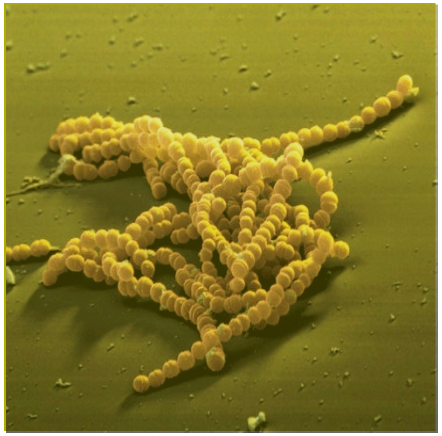
- Recombination was the major contributor of adaptive diversification (virulence factors, antibiotic resistance, niche adaptation)
- ST131 diverged some time prior 2000 into three closely related sublineages but is not strictly clonal

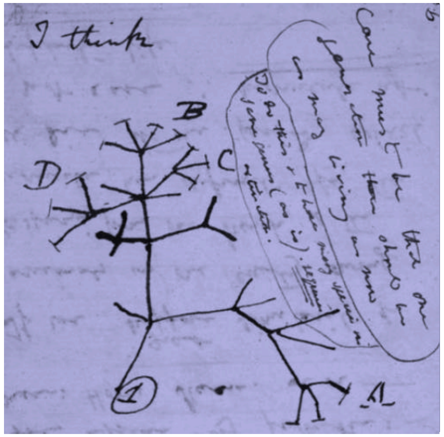
- What triggers higher recombinogenic potential? Antibiotic pressure?
- What makes ST131 so successful at colonising the urinary tract and other clinical sites?

Conclusions

- Importance of recombination in the evolution of bacteria
  - Functional
  - Evolutionary
- Multiple methods to detect recombination exist, choosing the right one is dependent on the dataset
  - Relatedness of the strains
  - Size of the dataset
- Huge increase of sequencing data available for bacterial populations to evaluate the impact of recombination more accurately

- **Beatson Lab**
  - **Scott Beatson**
  - Brian Forde
  - Mitchell Stanton-Cook
  - Elizabeth Skippington
- **SCMB/AID**
  - **Mark Schembri**
  - Makrina Totsika
  - Minh Duy Phan
  - Danilo Moriel
  - Kate Peters
- **UTS / i3 Institute**
  - Nico Petty
- **UQCCR**
  - David Paterson
  - Benjamin Rogers
- **Plymouth University**
  - Matt Upton
- **Cardiff University**
  - Tim Walsh
- **University of Calgary**
  - Johann Pitout
- **Universidad de Sevilla**
  - Jesus Rodriguez-Baño
  - Alvaro Pascual
- **Sanger Institute**
  - Gordon Dougan





AID | Australian Infectious Diseases Research Centre

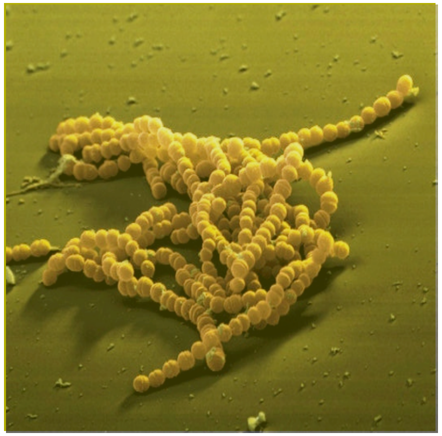
Australian Government
Australian Research Council

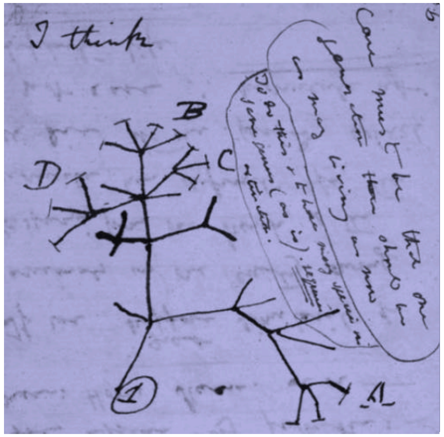
Australian Government
National Health and Medical Research Council

Nouri BEN ZAKOUR
@genomiss

SCMB – AID – University of Queensland

Thanks





Beatson Microbial Genomics Lab

Research

Nouri BEN ZAKOUR

@genomiss

SCMB – AID – University of Queensland