Mobilize your genome: Horizontal transfers, mobile elements and genome evolution in bacterial and eukaryotic pathogens

Molecular Epidemiology of Infectious Diseases

Lecture 8

March 2nd, 2026

Classic population genetic models of evolution focus on changes in genotype frequencies

Evolutionary Theory 101: Selection

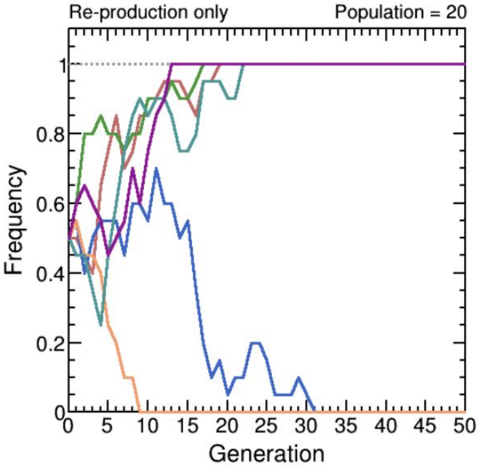
Selection “acts” on genetic variation (i.e. mutations or alleles) at one or a few loci to increase the frequency of higher fitness variants.

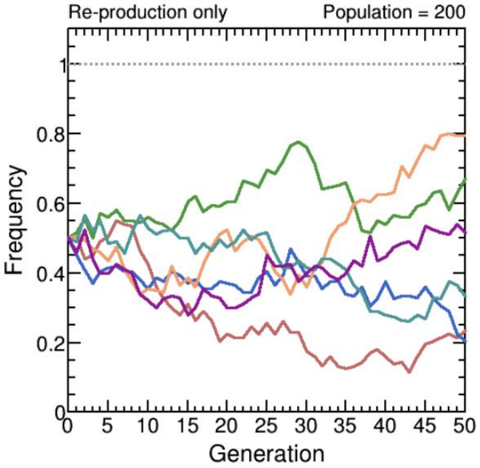
Selection can purge deleterious mutations and increase the frequency of beneficial mutations.

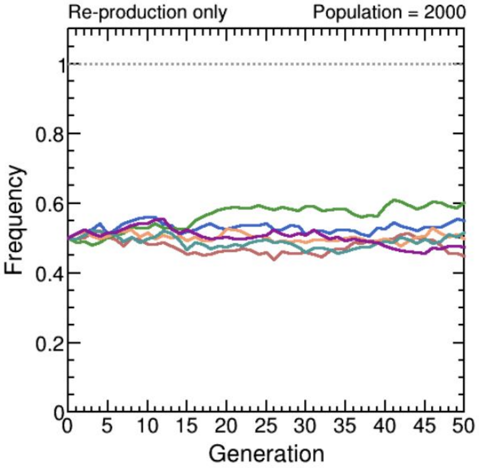
The strength of selection will depend on the fitness effects of mutations and the size of populations due to genetic drift.

Genetic drift

Genetic drift refers to stochastic fluctuations in genotype frequencies caused by random variation in reproduction and survival. Drift plays a larger role in smaller populations.





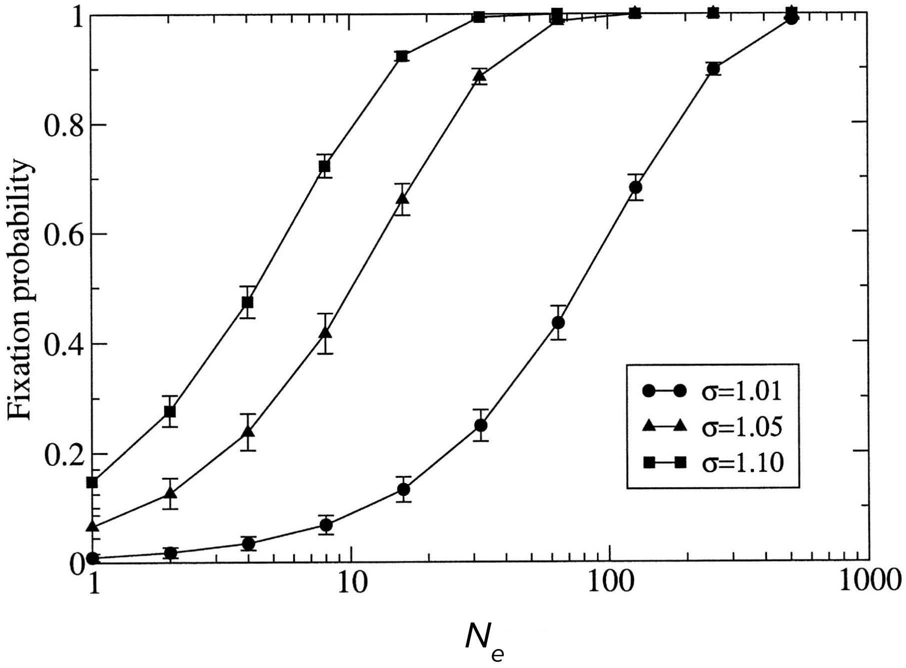


https://boundino.github.io/S188592web/drift.html

Genetic drift

The probability that a beneficial mutation reaches fixation (freq → 1.0) depends both on its selective advantage (s or σ) and the effective population size (N_e) – the number of individuals that contribute progeny to the next generation.

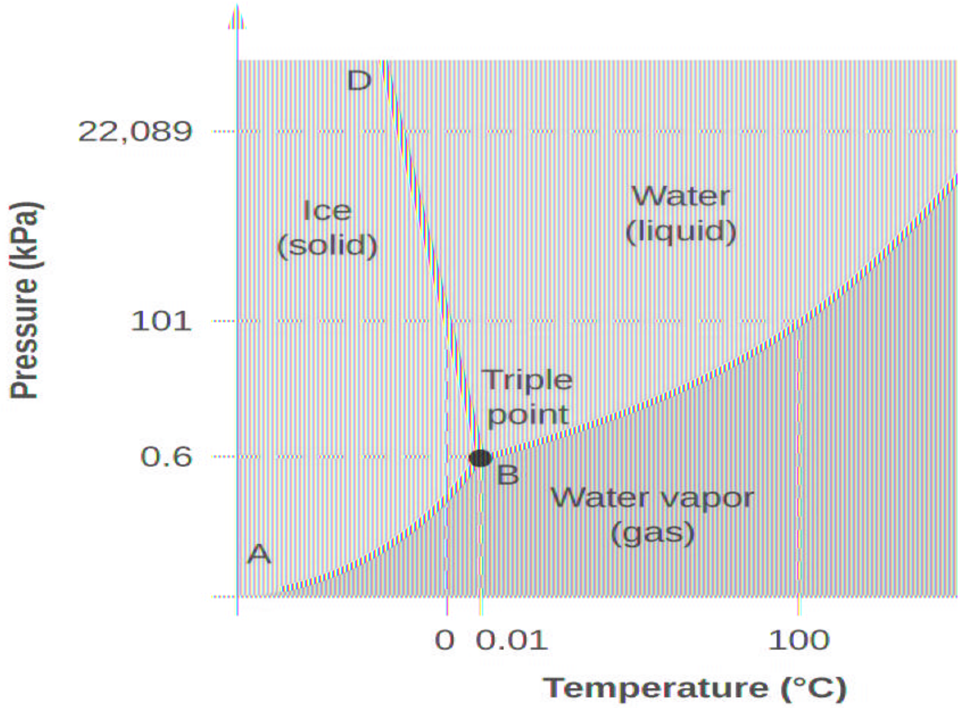
$s = w_{mut} - w_{wt}$



Wilke (Genetics, 2003)

Evolutionary phase diagrams

Phase diagrams describe how the large-scale properties of systems depend on key variables and identify ***critical points*** at which the qualitative behaviour of the system changes.

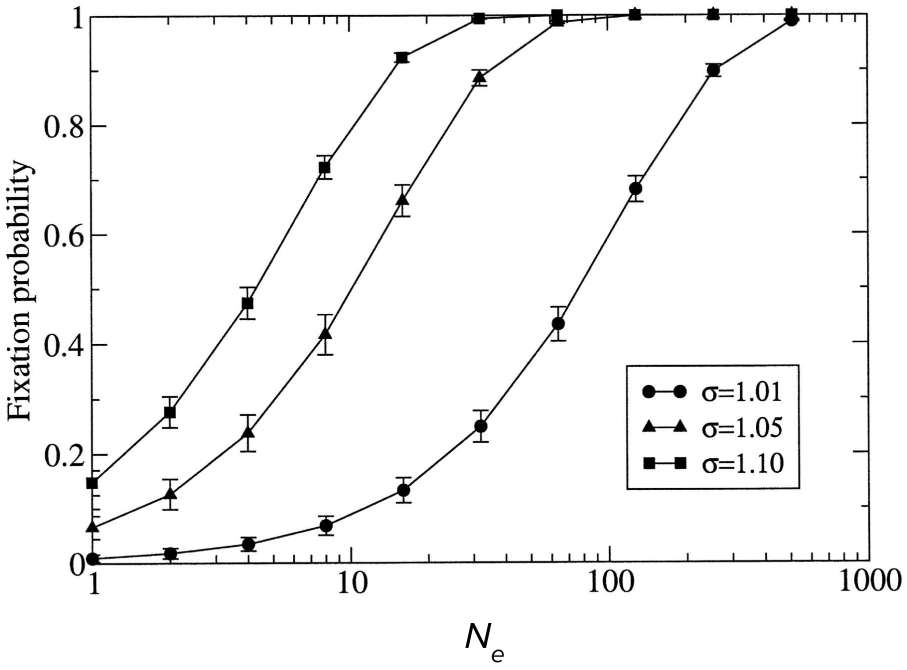


https://cnx.org/contents/oFoO44pW@5/Phase-Diagrams

Genetic drift

The probability that a beneficial mutation reaches fixation (freq → 1.0) depends both on its selective advantage (s or σ) and the effective population size (N_e) – the number of individuals that contribute progeny to the next generation.

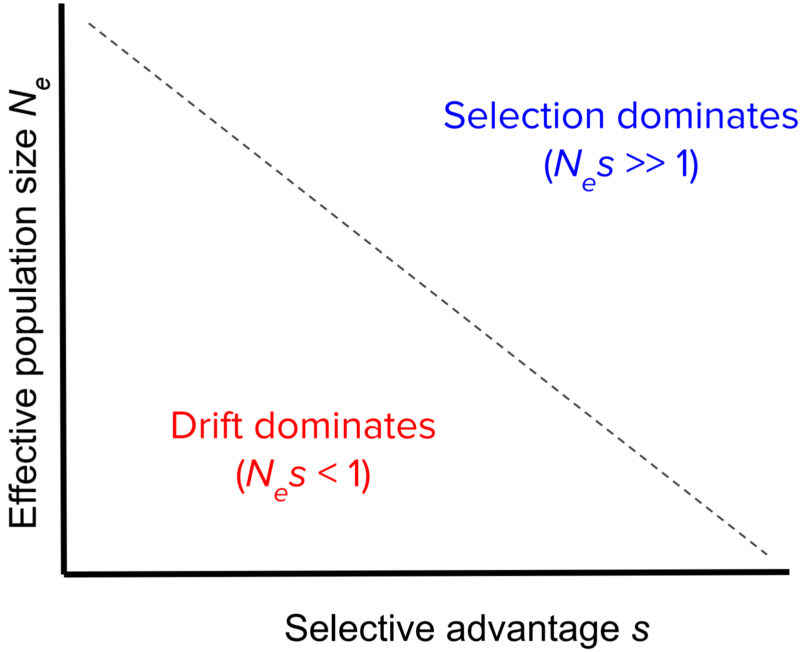
$$s = w_{mut} - w_{wt}$$



Wilke (Genetics, 2003)

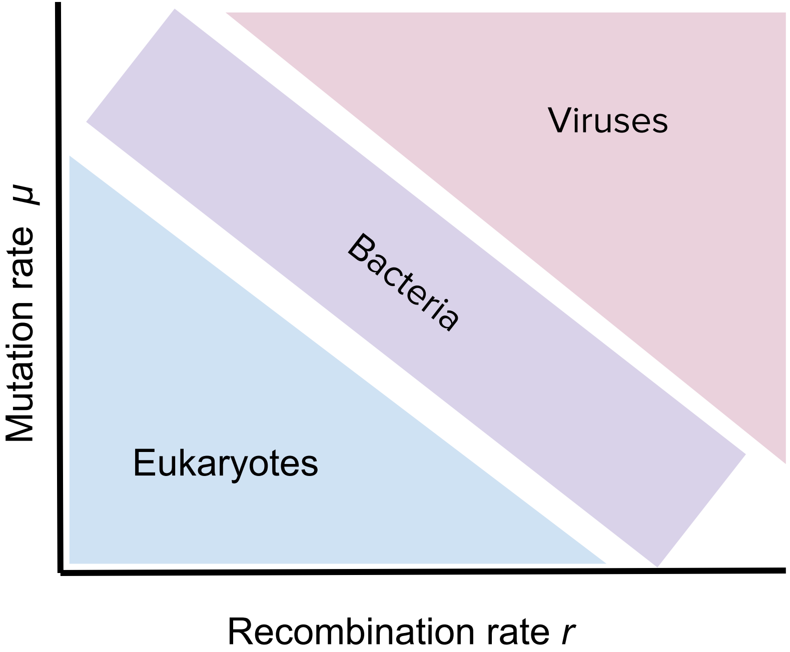
Selection vs. drift

The relative importance of selection versus drift is determined by $N_e s$



Ok, but how do pathogen genomes *actually* evolve?

Recombination vs. mutation



Recombination vs. mutation rates

The ratio r/m varies widely among different microbial pathogens

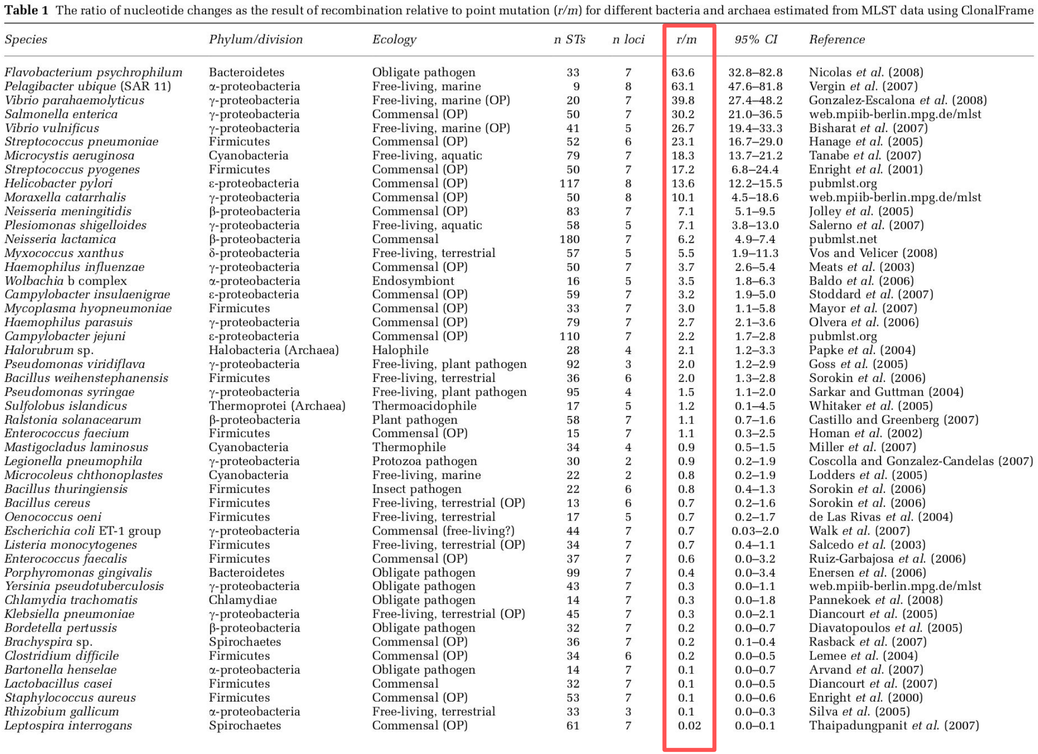


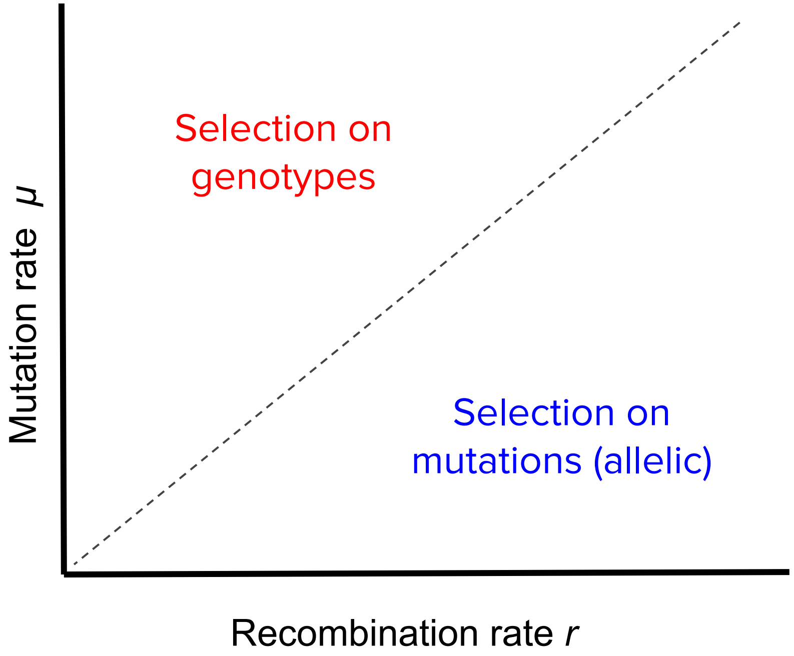
Table 1 The ratio of nucleotide changes as the result of recombination relative to point mutation (*r/m*) for different bacteria and archaea estimated from MLST data using ClonalFrame

Species	*Phylum/division*	*Ecology*	*n STs*	*n loci*	*r/m*	*95% CI*	*Reference*
Flavobacterium psychrophilum	Bacteroidetes	Obligate pathogen	33	7	63.6	32.8–82.8	Nicolas *et al.* (2008)
Pelagibacter ubique (SAR 11)	α-proteobacteria	Free-living, marine	9	8	63.1	47.6–81.8	Vergin *et al.* (2007)
Vibrio parahaemolyticus	γ-proteobacteria	Free-living, marine (OP)	20	7	39.8	27.4–48.2	Gonzalez-Escalona *et al.* (2008)
Salmonella enterica	γ-proteobacteria	Commensal (OP)	50	7	30.2	21.0–36.5	web.mpiib-berlin.mpg.de/mlst
Vibrio vulnificus	γ-proteobacteria	Free-living, marine (OP)	41	5	26.7	19.4–33.3	Bisharat *et al.* (2007)
Streptococcus pneumoniae	Firmicutes	Commensal (OP)	52	6	23.1	16.7–29.0	Hanage *et al.* (2005)
Microcystis aeruginosa	Cyanobacteria	Free-living, aquatic	79	7	18.3	13.7–21.2	Tanabe *et al.* (2007)
Streptococcus pyogenes	Firmicutes	Commensal (OP)	50	7	17.2	6.8–24.4	Enright *et al.* (2001)
Helicobacter pylori	ε-proteobacteria	Commensal (OP)	117	8	13.6	12.2–15.5	pubmlst.org
Moraxella catarrhalis	γ-proteobacteria	Commensal (OP)	50	8	10.1	4.5–18.6	web.mpiib-berlin.mpg.de/mlst
Neisseria meningitidis	β-proteobacteria	Commensal (OP)	83	7	7.1	5.1–9.5	Jolley *et al.* (2005)
Plesiomonas shigelloides	γ-proteobacteria	Free-living, aquatic	58	5	7.1	3.8–13.0	Salerno *et al.* (2007)
Neisseria lactamica	β-proteobacteria	Commensal	180	7	6.2	4.9–7.4	pubmlst.net
Myxococcus xanthus	δ-proteobacteria	Free-living, terrestrial	57	5	5.5	1.9–11.3	Vos and Velicer (2008)
Haemophilus influenzae	γ-proteobacteria	Commensal (OP)	50	7	3.7	2.6–5.4	Meats *et al.* (2003)
Wolbachia b complex	α-proteobacteria	Endosymbiont	16	5	3.5	1.8–6.3	Baldo *et al.* (2006)
Campylobacter insulaenigrae	ε-proteobacteria	Commensal (OP)	59	7	3.2	1.9–5.0	Stoddard *et al.* (2007)
Mycoplasma hyopneumoniae	Firmicutes	Commensal (OP)	33	7	3.0	1.1–5.8	Mayor *et al.* (2007)
Haemophilus parasuis	γ-proteobacteria	Commensal (OP)	79	7	2.7	2.1–3.6	Olvera *et al.* (2006)
Campylobacter jejuni	ε-proteobacteria	Commensal (OP)	110	7	2.2	1.7–2.8	pubmlst.org
Halorubrum sp.	Halobacteria (Archaea)	Halophile	28	4	2.1	1.2–3.3	Papke *et al.* (2004)
Pseudomonas viridiflava	γ-proteobacteria	Free-living, plant pathogen	92	3	2.0	1.2–2.9	Goss *et al.* (2005)
Bacillus weihenstephanensis	Firmicutes	Free-living, terrestrial	36	6	2.0	1.3–2.8	Sorokin *et al.* (2006)
Pseudomonas syringae	γ-proteobacteria	Free-living, plant pathogen	95	4	1.5	1.1–2.0	Sarkar and Guttman (2004)
Sulfolobus islandicus	Thermoprotei (Archaea)	Thermoacidophile	17	5	1.2	0.1–4.5	Whitaker *et al.* (2005)
Ralstonia solanacearum	β-proteobacteria	Plant pathogen	58	7	1.1	0.7–1.6	Castillo and Greenberg (2007)
Enterococcus faecium	Firmicutes	Commensal (OP)	15	7	1.1	0.3–2.5	Homan *et al.* (2002)
Mastigocladus laminosus	Cyanobacteria	Thermophile	34	4	0.9	0.5–1.5	Miller *et al.* (2007)
Legionella pneumophila	γ-proteobacteria	Protozoa pathogen	30	2	0.9	0.2–1.9	Coscolla and Gonzalez-Candelas (2007)
Microcoleus chthonoplastes	Cyanobacteria	Free-living, marine	22	2	0.8	0.2–1.9	Lodders *et al.* (2005)
Bacillus thuringiensis	Firmicutes	Insect pathogen	22	6	0.8	0.4–1.3	Sorokin *et al.* (2006)
Bacillus cereus	Firmicutes	Free-living, terrestrial (OP)	13	6	0.7	0.2–1.6	Sorokin *et al.* (2006)
Oenococcus oeni	Firmicutes	Free-living, terrestrial	17	5	0.7	0.2–1.7	de Las Rivas *et al.* (2004)
Escherichia coli ET-1 group	γ-proteobacteria	Commensal (free-living?)	44	7	0.7	0.03–2.0	Walk *et al.* (2007)
Listeria monocytogenes	Firmicutes	Free-living, terrestrial (OP)	34	7	0.7	0.4–1.1	Salcedo *et al.* (2003)
Enterococcus faecalis	Firmicutes	Commensal (OP)	37	7	0.6	0.0–3.2	Ruiz-Garbajosa *et al.* (2006)
Porphyromonas gingivalis	Bacteroidetes	Obligate pathogen	99	7	0.4	0.0–3.4	Enersen *et al.* (2006)
Yersinia pseudotuberculosis	γ-proteobacteria	Obligate pathogen	43	7	0.3	0.0–1.1	web.mpiib-berlin.mpg.de/mlst
Chlamydia trachomatis	Chlamydiae	Obligate pathogen	14	7	0.3	0.0–1.8	Pannekoek *et al.* (2008)
Klebsiella pneumoniae	γ-proteobacteria	Free-living, terrestrial (OP)	45	7	0.3	0.0–2.1	Diancourt *et al.* (2005)
Bordetella pertussis	β-proteobacteria	Obligate pathogen	32	7	0.2	0.0–0.7	Diavatopoulos *et al.* (2005)
Brachyspira sp.	Spirochaetes	Commensal (OP)	36	7	0.2	0.1–0.4	Rasback *et al.* (2007)
Clostridium difficile	Firmicutes	Commensal (OP)	34	6	0.2	0.0–0.5	Lemee *et al.* (2004)
Bartonella henselae	α-proteobacteria	Obligate pathogen	14	7	0.1	0.0–0.7	Arvand *et al.* (2007)
Lactobacillus casei	Firmicutes	Commensal	32	7	0.1	0.0–0.5	Diancourt *et al.* (2007)
Staphylococcus aureus	Firmicutes	Commensal (OP)	53	7	0.1	0.0–0.6	Enright *et al.* (2000)
Rhizobium gallicum	α-proteobacteria	Free-living, terrestrial	33	3	0.1	0.0–0.3	Silva *et al.* (2005)
Leptospira interrogans	Spirochaetes	Commensal (OP)	61	7	0.02	0.0–0.1	Thaipadungpanit *et al.* (2007)

Vos & Didelot (ISME, 2008)

Recombination vs. mutation rates

The relative ratio of recombination versus mutation rates determines whether selection acts primarily on individual mutations or whole genotypes/haplotypes.



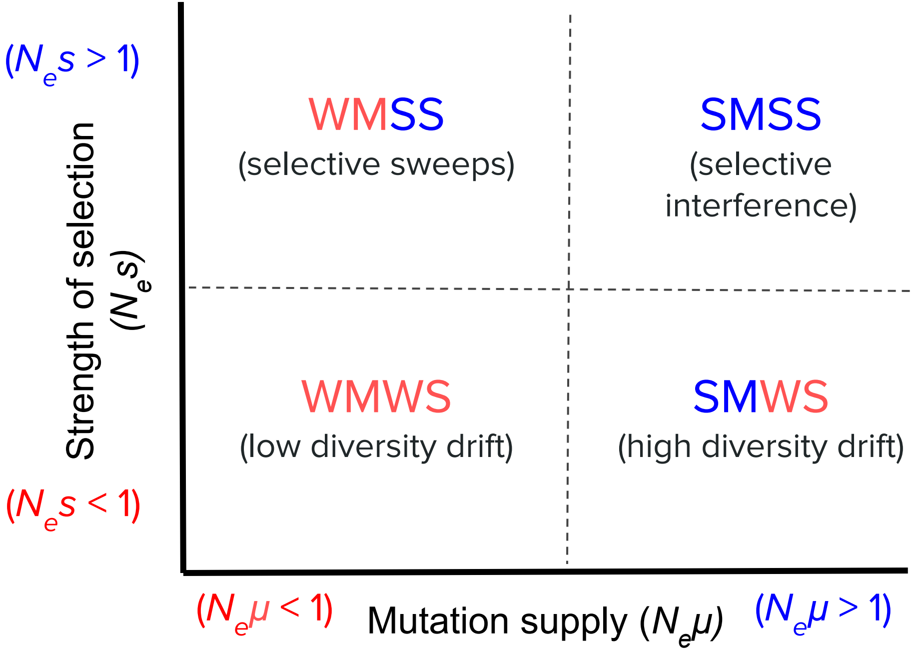
Selection on linked sites

When **$r < \mu$** mutations will be linked to other mutations in the same genetic background (haplotype). The fate of a mutation therefore depends on the fitness effects of other linked mutations (background or interference selection).

In contrast, when **$r > \mu$** recombination will reshuffle mutations onto different genetic backgrounds such that selection can act on individual mutations (allelic selection).

The strength of selection vs. mutation

The strength of selection ($N_e s$) versus the mutational supply ($N_e \mu$) determines whether mutations co-occur and compete with one another in a population.



($N_e s > 1$)

Strength of selection ($N_e s$)

($N_e s < 1$)

WMSS
(selective sweeps)

SMSS
(selective interference)

WMWS
(low diversity drift)

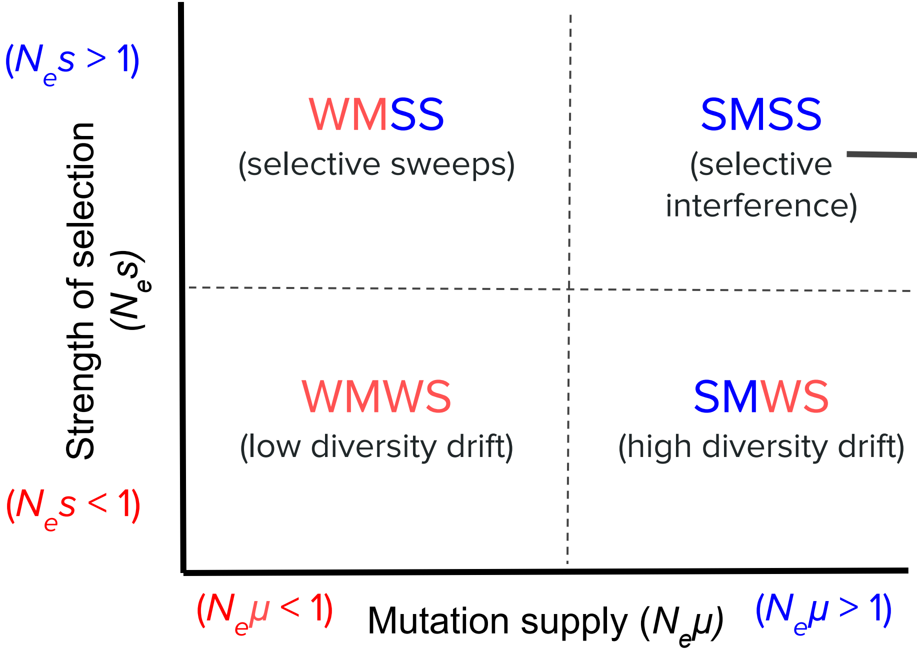
SMWS
(high diversity drift)

($N_e \mu < 1$) Mutation supply ($N_e \mu$) ($N_e \mu > 1$)

Sneigowski & Gerrish (Phil. Trans. Roy. Soc. B, 2010)

The strength of selection vs. mutation

The strength of selection ($N_e s$) versus the mutational supply ($N_e \mu$) determines whether mutations co-occur and compete with one another in a population.



$(N_e s > 1)$

Strength of selection ($N_e s$)

$(N_e s < 1)$

WMSS
(selective sweeps)

SMSS
(selective interference)

WMWS
(low diversity drift)

SMWS
(high diversity drift)

$(N_e \mu < 1)$ Mutation supply ($N_e \mu$) $(N_e \mu > 1)$

Leads to “clonal interference” regime where a lot of traditional pop gen theory breaks down

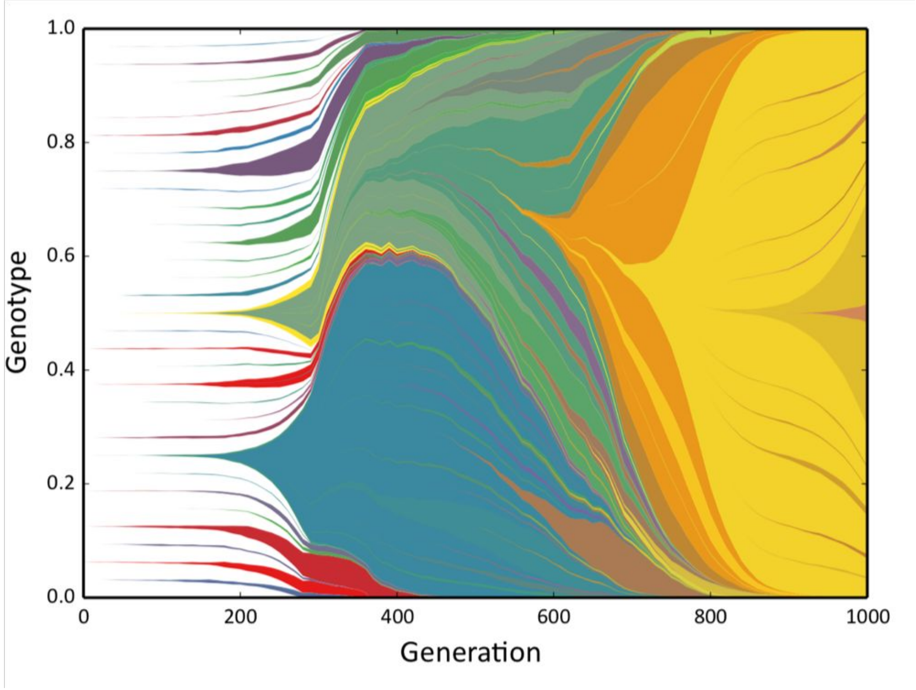
Sneigowski & Gerrish (Phil. Trans. Roy. Soc. B, 2010)

Clonal interference

Clonal interference arises in populations with high mutation rates and large population sizes.

Multiple lineages with beneficial mutations compete with one another.

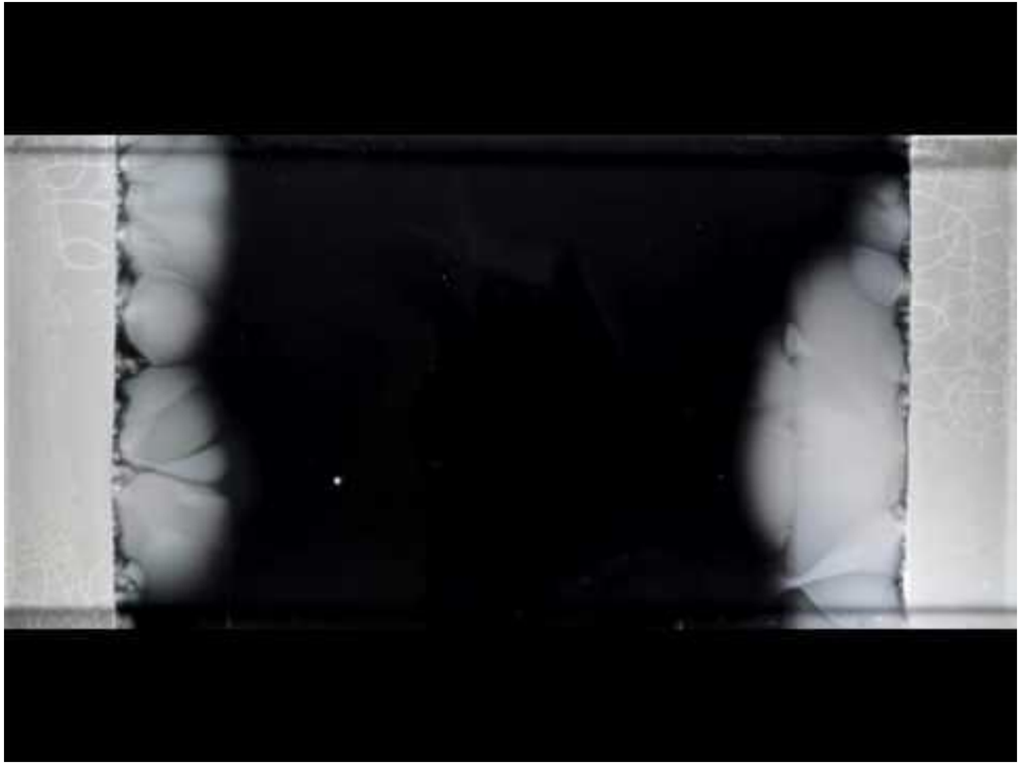
Quite different from “typical” pop gen scenarios where each mutant arises in temporal succession.



Cvijovic et al. (Trends in Genetics, 2018)

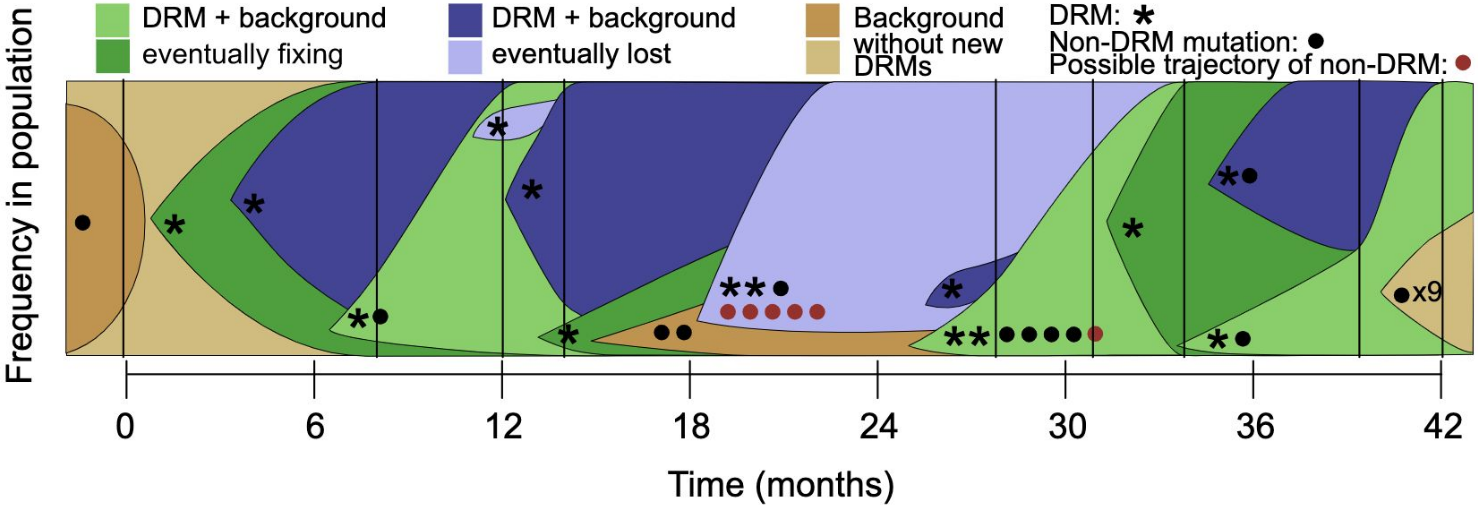
Clonal interference through competition

The Microbial Evolution and Growth Arena (MEGA) experiments show multiple strains of antibiotic resistant *E. coli* competing with one other for space.



Clonal interference in *M. tuberculosis*

Dynamics of TB clones within a host treated with a series of antibiotics. Only 7 out of 12 beneficial DRM mutants reach fixation while the rest are outcompeted.



Eldhom et al. (Genome Bio, 2014)

Clonal interference: summary

Role of genetic drift becomes negligible as competition creates strong selection for highly fit genotypes.

Increases chance that “best” genotype with the largest fitness advantage goes to fixation. This genotype may often carry multiple beneficial mutations.

However, interference can actually slow down the rate of adaptation (increases in mean pop fitness) as multiple beneficial genotypes will compete against one another.

How do multi-drug resistant bacteria evolve?

Resistance evolution in MRSA

Methicillin resistant *Staphylococcus aureus* first emerged in the 1960’s and causes dangerous bloodstream infections with roughly 30% mortality.

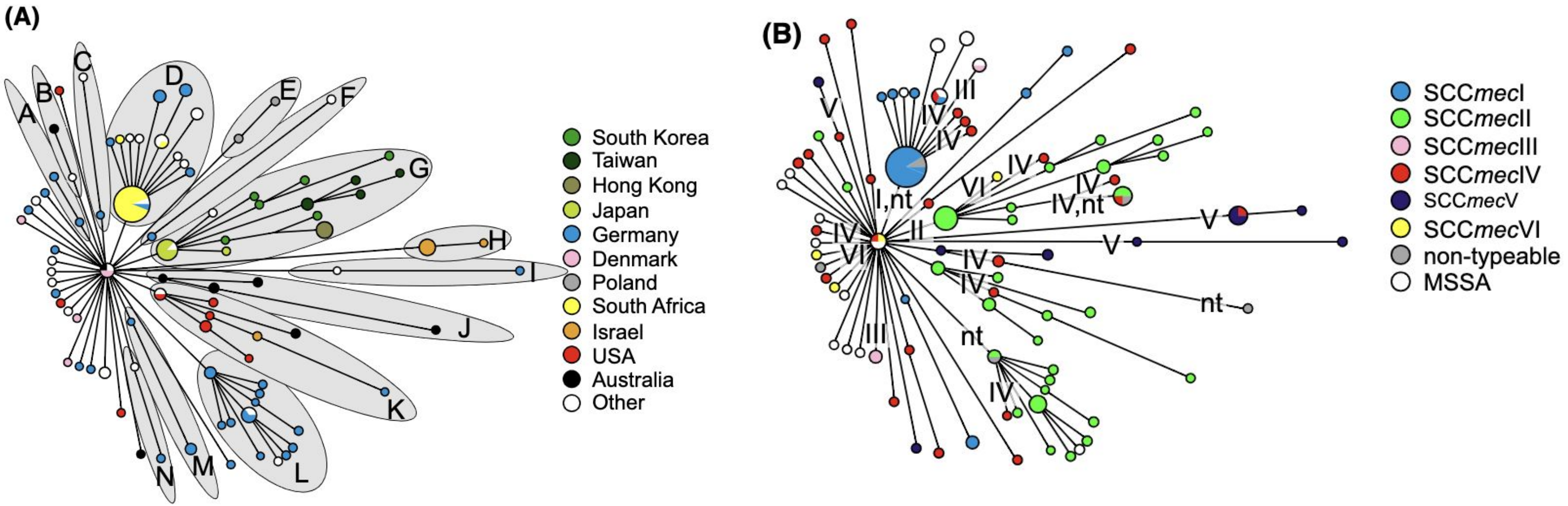
S. aureus is highly clonal with a low recombination rate and only a few lineages are responsible for most infections.

Methicillin resistance is acquired via the staphylococcal cassette chromosome *SCCmec,* a mobile genetic element that integrates a cassette of genes into the bacterial chromosome (most likely by transduction from bacteriophages).

SCCmec acquisition is a classic example of horizontal gene transfer (HGT).

Acquisition of methicillin resistance

S. aureus phylogenies show independent acquisitions of *SCCmec* cassettes across different lineages suggesting frequent horizontal transfers.



Nübel *et al.* (PNAS, 2008)

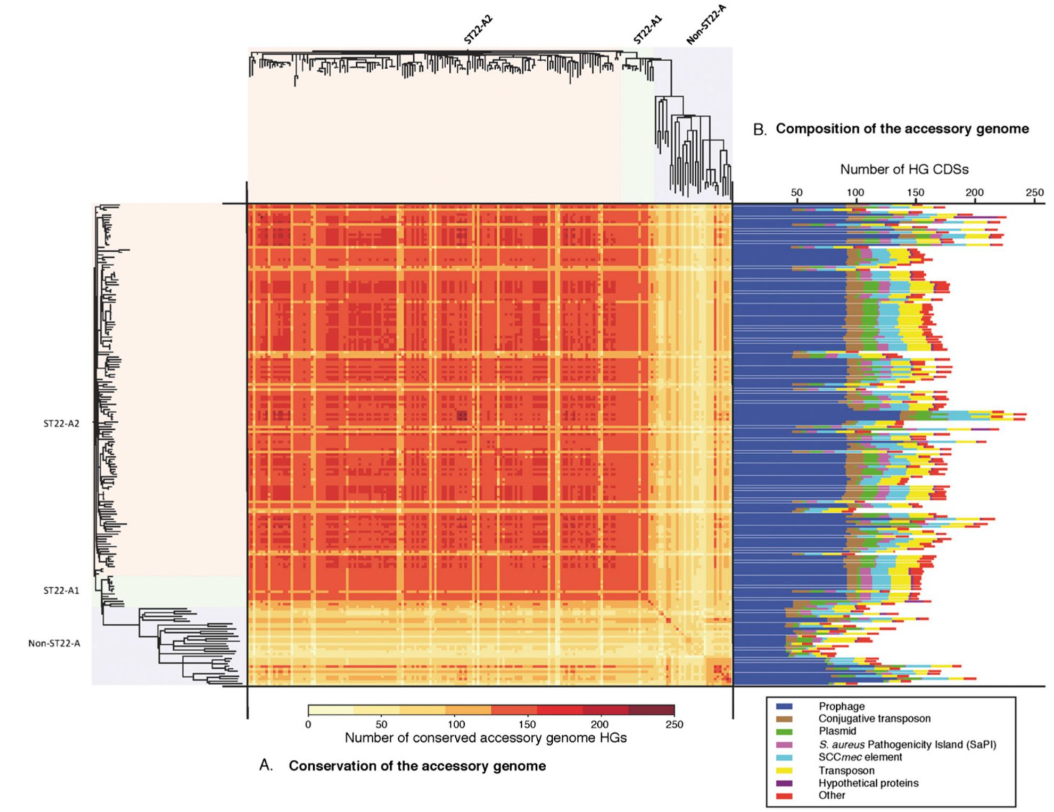
The pangenome concept

The **core genome** contains the highly conserved genes with often essential functions (i.e. DNA replication) that are present within all members of a species.

The **accessory genome** includes gene content that varies between strains including chromosomal cassettes, prophages, transposons and pathogenicity islands. These genes can be chromosomal or extrachromosomal (e.g. on plasmids).

Prokaryote populations are often characterized by extensive sharing of accessory genes due to horizontal gene transfers and subsequent gene losses.

The MRSA accessory genome



Holden *et al.* (Genome Research, 2013)

Evolution of multidrug resistant MRSA

Increased fitness of initially methicillin resistance strains leads to large clonal expansions.

Accumulation of additional genes conferring resistance to other antibiotics occurred through further HGTs, leading to superfit multidrug resistance clones.

These multi-drug resistant (MDR) strains have acquired resistance to several major classes of antibiotics including fluoroquinolones and erythromycin.

MDR MRSA

“MRSA epitomizes a now all-too-familiar evolutionary route by which successful AMR clones emerge in response to local antimicrobial usage, undergo population expansion under selection from sustained antimicrobial exposure and then explode into pandemic spread”

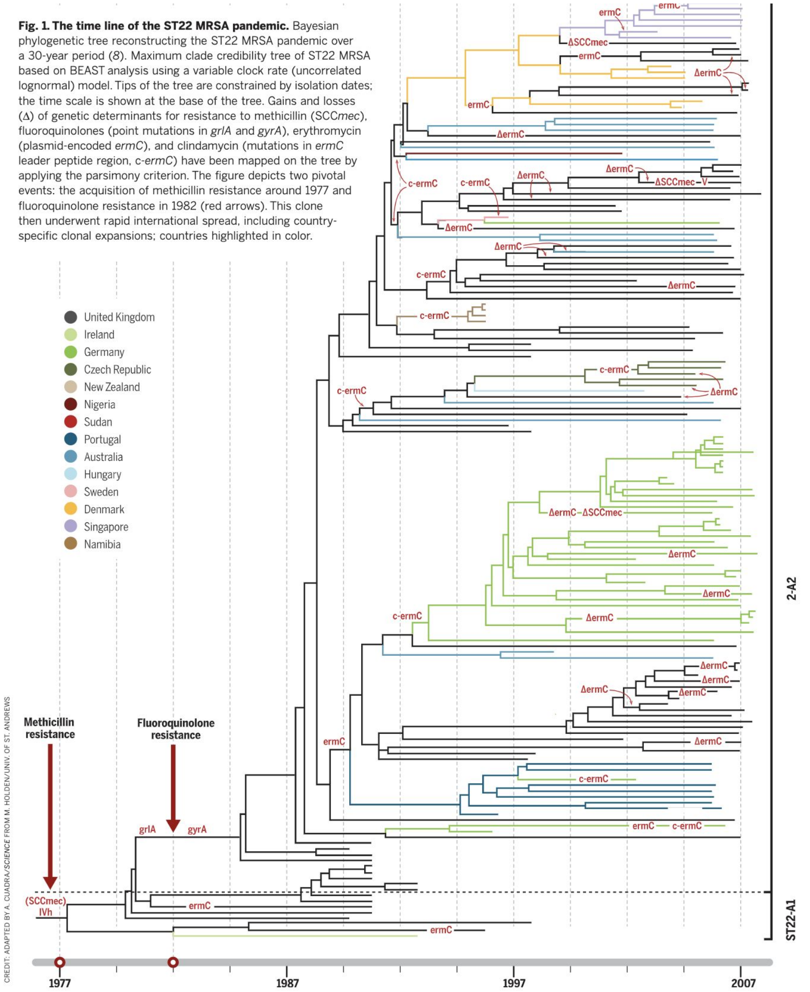


Fig. 1. The time line of the ST22 MRSA pandemic. Bayesian phylogenetic tree reconstructing the ST22 MRSA pandemic over a 30-year period (*8*). Maximum clade credibility tree of ST22 MRSA based on BEAST analysis using a variable clock rate (uncorrelated lognormal) model. Tips of the tree are constrained by isolation dates; the time scale is shown at the base of the tree. Gains and losses (Δ) of genetic determinants for resistance to methicillin (SCC*mec*), fluoroquinolones (point mutations in *grlA* and *gyrA*), erythromycin (plasmid-encoded *ermC*), and clindamycin (mutations in *ermC* leader peptide region, c-*ermC*) have been mapped on the tree by applying the parsimony criterion. The figure depicts two pivotal events: the acquisition of methicillin resistance around 1977 and fluoroquinolone resistance in 1982 (red arrows). This clone then underwent rapid international spread, including country-specific clonal expansions; countries highlighted in color.

CREDIT: ADAPTED BY A. CUADRA/*SCIENCE* FROM M. HOLDEN/UNIV. OF ST. ANDREWS

Baker *et al.* (Science, 2017)

“Two-speed” genomes in eukaryotes

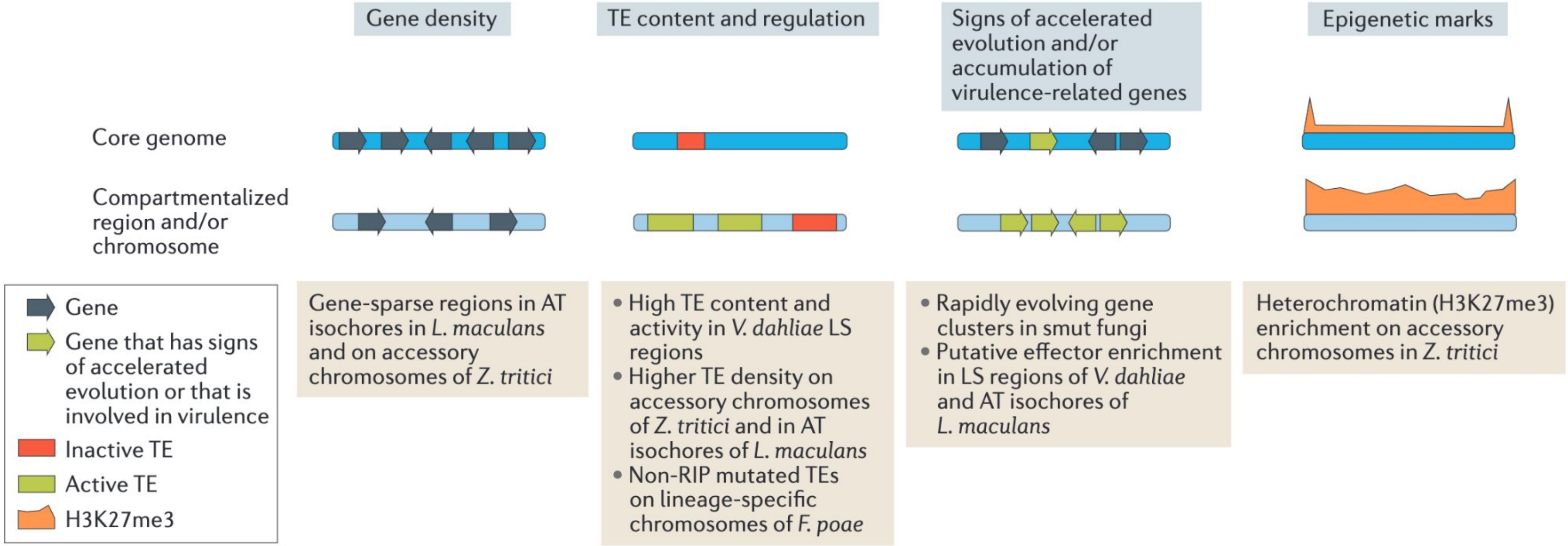
Pathogenic fungi often have virulence-related and effector genes involved in host-adaptation compartmentalized to genomic regions with elevated mutations rates and high densities of mobile genetic elements, including:

Accessory chromosomes: lineage-specific chromosomes with low gene density but high mutation rates that can move horizontally between lineages.

TE-rich compartments: regions of the core genome that are gene-sparse but highly variable due to transposable elements or other mobile genetic elements.

Dong *et al.* (2015)

Fast evolving genome compartments



Gene density

TE content and regulation

Signs of accelerated evolution and/or accumulation of virulence-related genes

Epigenetic marks

Core genome

Compartmentalized region and/or chromosome

- Gene
- Gene that has signs of accelerated evolution or that is involved in virulence
- 🟥 Inactive TE
- 🟩 Active TE
- H3K27me3

Gene-sparse regions in AT isochores in *L. maculans* and on accessory chromosomes of *Z. tritici*

- High TE content and activity in *V. dahliae* LS regions
- Higher TE density on accessory chromosomes of *Z. tritici* and in AT isochores of *L. maculans*
- Non-RIP mutated TEs on lineage-specific chromosomes of *F. poae*

- Rapidly evolving gene clusters in smut fungi
- Putative effector enrichment in LS regions of *V. dahliae* and AT isochores of *L. maculans*

Heterochromatin (H3K27me3) enrichment on accessory chromosomes in *Z. tritici*

Möller and Stukenbrock (Nature Rev. Micro., 2017)

Recombine often or perish

Moderately high mutation rates but low recombination rates cause competition and clonal interference between high-fitness lineages.

Rapid adaptation to selective pressures like antimicrobials occurs by frequent horizontal transfers of beneficial genetic elements in the accessory genome.

Successful lineages therefore tend to be clones that have acquired multiple beneficial genes or other genetic elements through HGT.

How clonal are bacteria and other microbial pathogens?

Clonal vs. horizontal evolution

Clonal evolution occurs by vertical descent where genetic material is passed from parents to children.

In contrast, genetic material can be exchanged **horizontally** between lineages through recombination or horizontal gene transfers.

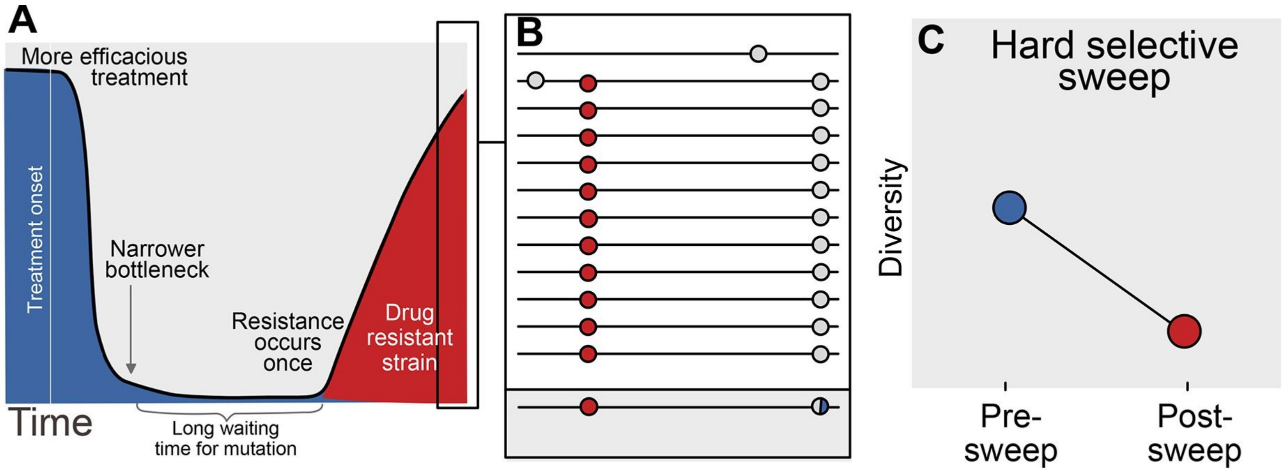
What determines clonality?

Rates of recombination and horizontal exchange have traditionally been thought of as the key determinants of clonality.

However, recombination also interacts with selection to determine how clonal a population is at any particular point in time.

Selective sweeps eliminate diversity

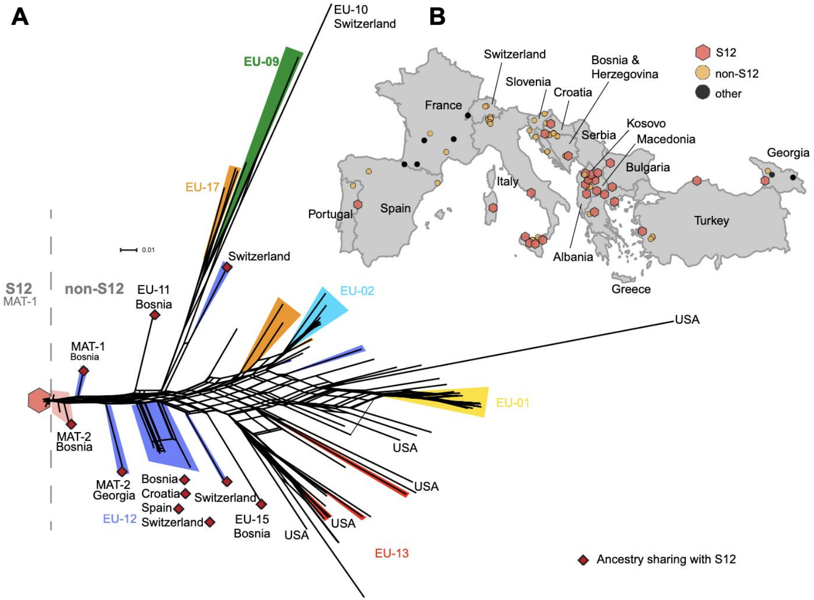
Strong selection can lead to rapid clonal expansions and genome-wide selective sweeps of linked variants.

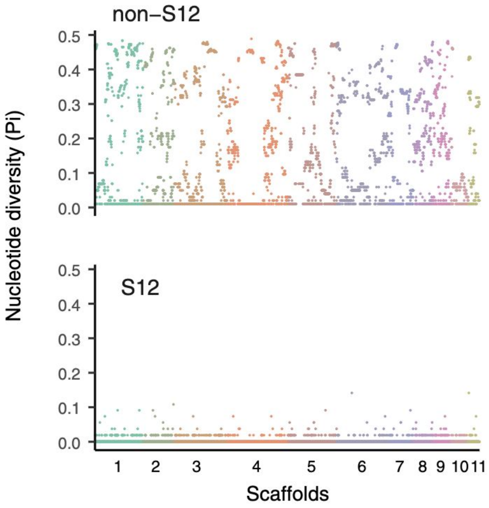


Feder *et al.* (eLife, 2016)

Clonal expansions of chestnut blight

The invasive S12 genotype is an asexual and more virulent lineage of *Cryphonectria parasitica* undergoing a clonal expansion in European chestnut trees.





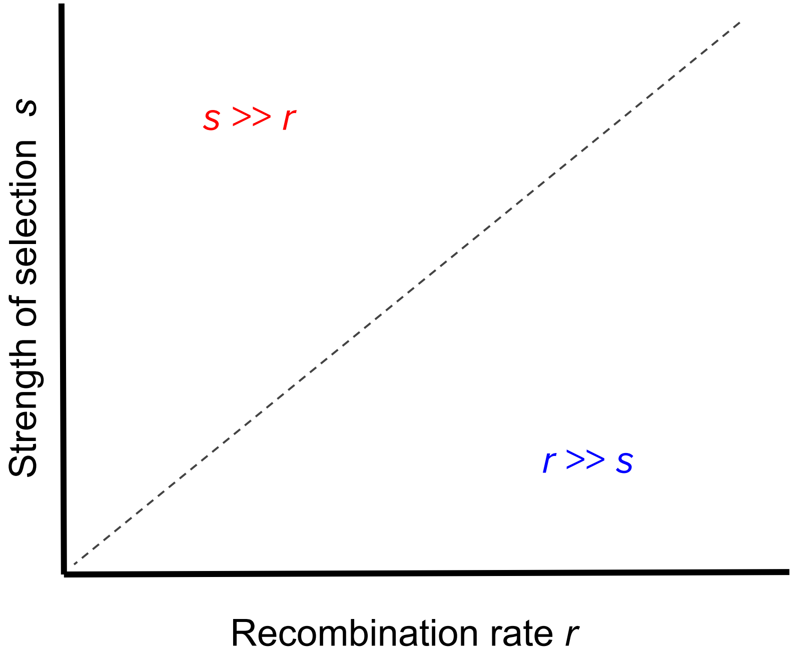
Stauber *et al.* (eLife, 2021)

Recombination vs. selection

The strength of selection relative to the recombination rate determines how selective sweeps impact diversity elsewhere in the genome.

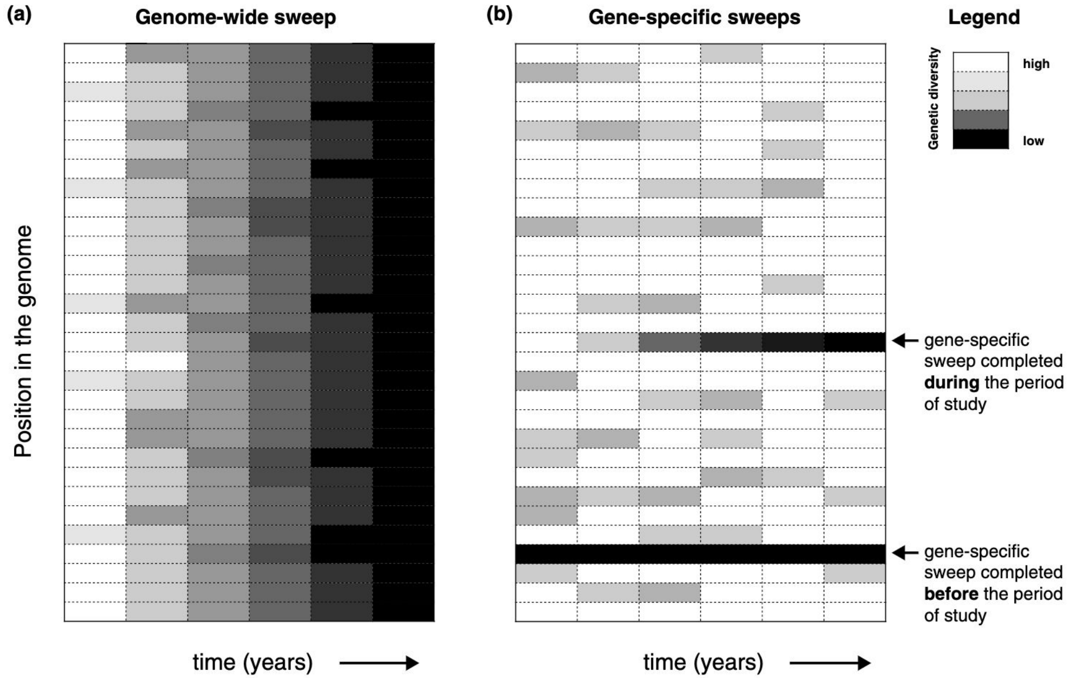
Genome-wide selective sweeps

$s = w_{mut} - w_{wt}$



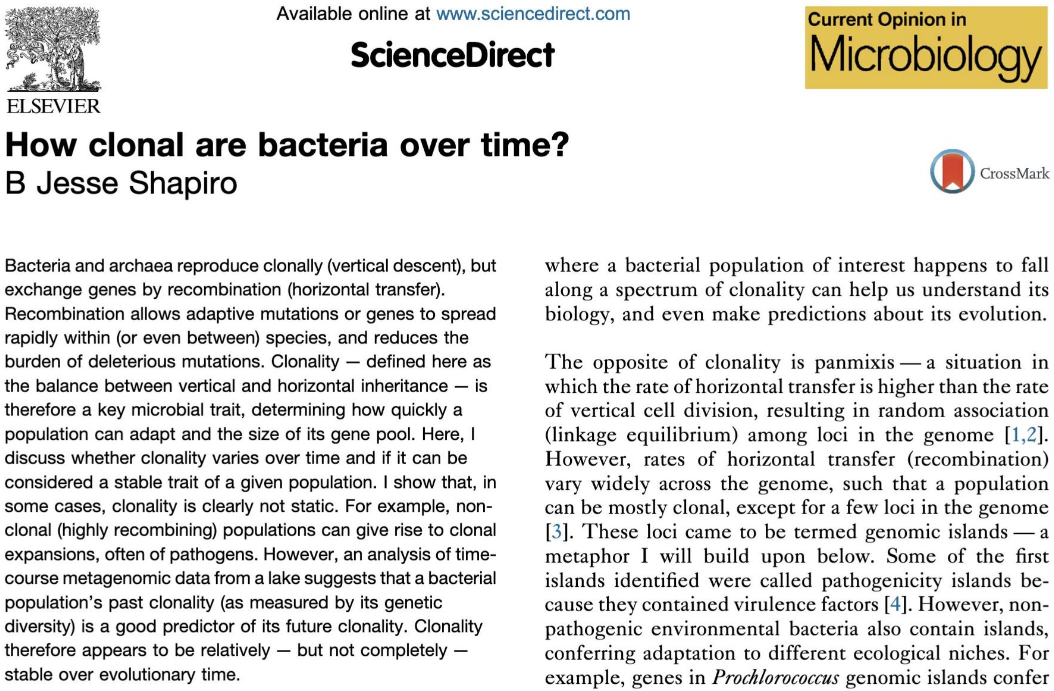
Gene-specific selective sweeps

Gene-specific vs. genome-wide sweeps



Shapiro (Curr. Op. Micro., 2016)

Recommended reading



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Microbiology

How clonal are bacteria over time?

B Jesse Shapiro

CrossMark

Bacteria and archaea reproduce clonally (vertical descent), but exchange genes by recombination (horizontal transfer). Recombination allows adaptive mutations or genes to spread rapidly within (or even between) species, and reduces the burden of deleterious mutations. Clonality — defined here as the balance between vertical and horizontal inheritance — is therefore a key microbial trait, determining how quickly a population can adapt and the size of its gene pool. Here, I discuss whether clonality varies over time and if it can be considered a stable trait of a given population. I show that, in some cases, clonality is clearly not static. For example, non-clonal (highly recombining) populations can give rise to clonal expansions, often of pathogens. However, an analysis of time-course metagenomic data from a lake suggests that a bacterial population's past clonality (as measured by its genetic diversity) is a good predictor of its future clonality. Clonality therefore appears to be relatively — but not completely — stable over evolutionary time.

where a bacterial population of interest happens to fall along a spectrum of clonality can help us understand its biology, and even make predictions about its evolution.

The opposite of clonality is panmixis — a situation in which the rate of horizontal transfer is higher than the rate of vertical cell division, resulting in random association (linkage equilibrium) among loci in the genome [1,2]. However, rates of horizontal transfer (recombination) vary widely across the genome, such that a population can be mostly clonal, except for a few loci in the genome [3]. These loci came to be termed genomic islands — a metaphor I will build upon below. Some of the first islands identified were called pathogenicity islands because they contained virulence factors [4]. However, non-pathogenic environmental bacteria also contain islands, conferring adaptation to different ecological niches. For example, genes in *Prochlorococcus* genomic islands confer

Shapiro (Curr. Op. Micro., 2016)

Clonality can dynamically vary over time

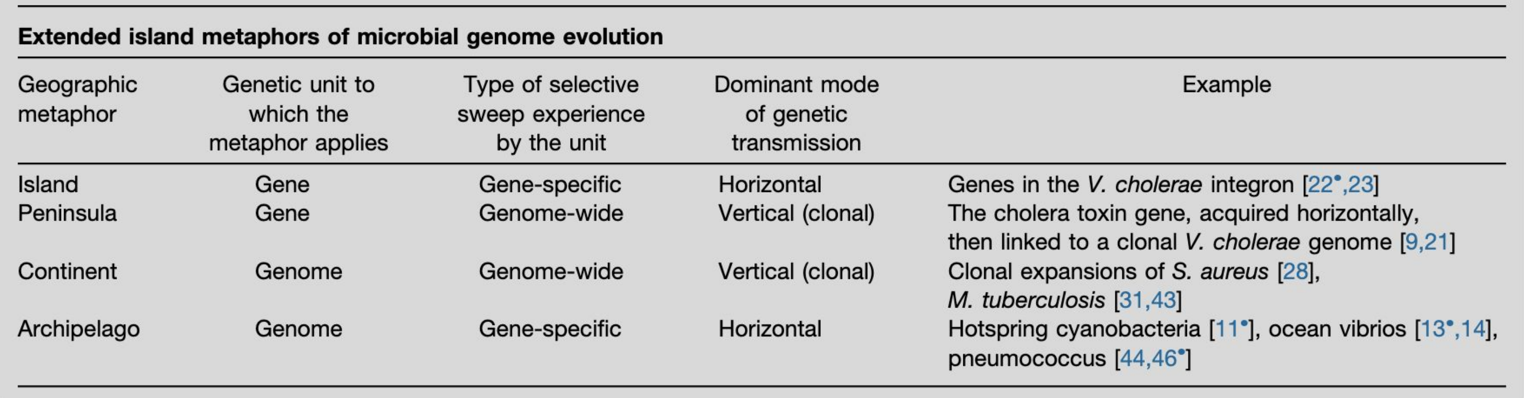
Genome-wide selective sweeps increase clonality even if recombination rates were historically high.

Genomic islands, pieces of DNA that were transferred horizontally, can therefore become peninsulas linked by the conserved regions of the genome (i.e. the continents).

Continents can be broken up over time into archipelagos by recombination.

The extended island metaphor

Clonality varies considerably between different species and may reflect their recent demographic and selective history.



Extended island metaphors of microbial genome evolution

Geographic metaphor	Genetic unit to which the metaphor applies	Type of selective sweep experience by the unit	Dominant mode of genetic transmission	Example
Island	Gene	Gene-specific	Horizontal	Genes in the *V. cholerae* integron [22•,23]
Peninsula	Gene	Genome-wide	Vertical (clonal)	The cholera toxin gene, acquired horizontally, then linked to a clonal *V. cholerae* genome [9,21]
Continent	Genome	Genome-wide	Vertical (clonal)	Clonal expansions of *S. aureus* [28], *M. tuberculosis* [31,43]
Archipelago	Genome	Gene-specific	Horizontal	Hotspring cyanobacteria [11•], ocean vibrios [13•,14], pneumococcus [44,46•]

Shapiro (Curr. Op. Micro., 2016)

Conclusions

The evolutionary dynamics of populations are shaped by the relative strength of mutation, drift, selection and recombination. Thinking in terms of phase diagrams can help us distinguish between qualitatively different regimes.

High mutation rates coupled with large pop sizes often lead to clonal interference where multiple variants simultaneously compete.

Horizontal gene transfers (HGT) allow for rapid adaptation (e.g. multi-drug resistance) even in otherwise clonal populations.

How clonal a population is at present depends not just on recombination but also on its demographic history including bouts of strong selection.