Title: How clonal are bacteria over time?

Author: B. Jesse Shapiro[1]

Affiliation

1. Département de sciences biologiques, Université de Montréal, Montréal, QC H3C 3J7, Canada

* jesse.shapiro@umontreal.ca

Abstract

Bacteria and archaea reproduce clonally (vertically), but exchange genes by recombination (horizontal transfer). Recombination allows adaptive mutations or genes to spread within (or between) species. Clonality – the balance between vertical and horizontal inheritance – is therefore a key microbial trait, determining how quickly a population can adapt. Here, I consider whether clonality can be considered a stable trait of a given population. In some cases, clonality changes over time: non-clonal (recombining) populations can give rise to clonal expansions. However, an analysis of time-course metagenomic data suggests that a bacterial population's past clonality is indicative of its future clonality. Thus, a population's evolutionary potential – whether it is likely to retain genetic diversity or not – can in principle be predicted from its past.

Introduction

Here, I revisit the question posed in the title of a classic paper by John Maynard Smith and colleagues [1]: How clonal are bacteria? More specifically how does clonality vary among different microbial populations and over time? First, what do we mean by clonality? Perfectly clonal bacteria replicate by cell division (vertical descent) and evolve by random mutations that occur during DNA replication. There is negligible horizontal transfer of DNA by recombination across the resulting tree of vertical descent. Very few (if any) natural bacterial populations fit this theoretical definition of clonality. Or, as discussed below, they might only fit it for a short amount of time. However, knowing where a bacterial population 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 [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. 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 [3]. However, non-pathogenic environmental bacteria also contain islands, conferring adaptation to different ecological niches. For example, genes in *Prochlorococcus* genomic islands confer adaptation to light and nutrient conditions [4,5]. But islands need not confer niche adaptation to their host genome; they can be neutral to host fitness or even detrimental, selfish parasites. Here, I define genomic islands broadly as any piece of DNA that is transferred horizontally (by either homologous or nonhomologous recombination) from cell to cell and therefore evolves independently (*i.e.* is unlinked) from the rest of the genome.

Are some islands really peninsulas?

47 In the classic analogy, an island is totally disconnected from the mainland, meaning that genes in
48 the island evolve independently of the genome (**Table 1**). Examples of islands that fit this strict
49 independence might include integrated phages and other "selfish" elements, or genes that reside
50 in a particular niche but not in a particular genome (*e.g.* a gene ecology model [6]). Peninsulas
51 provide an analogy that might better describe how islands are related to microbial genomes. A
52 peninsula (or "presque-ile," from the French for "almost island") is a geographic term for a very
53 narrow strip of land connected to (but distinct from) the mainland. An island is evolutionarily
54 independent of the mainland genome, but their fates may become linked, forming a peninsula. For
55 example, a bacterium may acquire a gene from a vast microbial gene pool. This gene allows the
56 bacterium to invade a new ecological niche, triggering a clonal expansion in which the fate of the
57 gene and its new host genome are linked, at least for the duration of the clonal expansion. One
58 such example could be *Yersinia pestis,* which acquired a single gene allowing flea-borne
59 transmission and triggering a clonal expansion in the form of Plague pandemics [7]. Another
60 peninsula, the prophage-encoded cholera toxin, and its links to the mainland *Vibrio cholerae*
61 genome [8,9], is discussed below.
62
63 **Are some genomes archipelagos?**
64 The very concept of one or a few islands implies a contrast with the large, clonal genomic
65 mainland or continent. But some microbial genomes may contain so many islands that there is no
66 mainland, only a vast archipelago (**Table 1**). A striking recent example is a population of
67 hotspring cyanobacteria in which virtually every gene in the genome evolved independently due
68 to frequent recombination [10], leading the authors to call the population "quasi-sexual"
69 (panmictic). The Asian ocean population of *Vibrio parahaemolyticus* also form a panmictic gene
70 pool, with each recombination block of ~1.8 kbp evolving independently [11]. Similarly, almost
71 every gene in a population of *Vibrio cyclitrophicus* genomes showed signals of recombination
72 over relatively recent time scales [12]. Such apparently high rates of recombination in natural

73 populations were mysterious at first, contradicting recombination rates measured in the lab
74 [13,14] and predicted by theory [15,16]. However, new models (discussed below) suggest
75 mechanisms capable of explaining such surprisingly panmictic populations [17,18].
76
77 **Clonal expansions from panmictic pools**
78 Archipelagos are not necessarily static over time; they can sometimes coalesce into continents.
79 Given the right ecological opportunity, a genome from a panmictic gene pool can escape the
80 "gravitational pull" of recombination and take off into a clonal expansion (resulting in “epidemic”
81 population structure [1]). An example mentioned earlier is *V. cholerae*, a genetically diverse
82 group of coastal marine bacteria, some of which cause cholera. Virulence is mainly determined
83 by two loci in the genome: the cholera toxin and the toxin-coregulated pilus. Both genes are
84 frequently gained and lost by recombination [19,20], but are always found in one lineage of *V.*
85 *cholerae* – the lineage causing severe disease with pandemic potential, known as the phylocore
86 genome (PG) group [9]. It remains a mystery why the PG lineage evolved once, and only once. If
87 PG *V. cholerae* really did evolve just once, this would be surprising because *V. cholerae* draws on
88 a diverse, global gene pool and can be considered panmictic [21]. Therefore multiple different
89 lineages would be expected to acquire the two (or perhaps a handful of) genetic elements required
90 for pandemic disease. This leads to the hypothesis that pandemic cholera emergence is *selection*
91 *limited* rather than *diversity limited*. In other words, benign *V. cholerae* strains constantly acquire
92 virulence genes. However, these strains rarely encounter the right ecological niche to flourish,
93 *e.g.* a human population consuming brackish water. "The right niche" has appeared a few times in
94 human history: for example in India in the 1800s, when the Classical lineage evolved, and again
95 in Indonesia in the 1950s, when the El Tor lineage evolved [22]. El Tor and Classical lineages
96 have different cholera toxin alleles, but both evolved from within the PG group, suggesting that
97 the PG genomic background is adapted to acquire and exploit virulence genes. When the “right”
98 conditions appear, the PG lineage, along with its virulence factors, takes off in a clonal

99 expansions which continue to wreak havoc today (*e.g.* cholera pandemics from the 1800s to
100 today, all caused by the PG clonal group). The virulence factors, previously islands in an
101 archipelago, became a peninsula connected to the PG mainland. The linkage between virulence
102 factors and PG remains imperfect because different variants of the cholera toxin continue to flow
103 in and out of the PG continent [9,19]; hence the toxin remains a peninsula, not firmly part of the
104 mainland.

105 *V. cholerae* is a particularly well-characterized example of a panmictic gene pool giving
106 rise to a clonal expansion, but similar evolutionary dynamics are seen in other pathogens as well.
107 For example, enterotoxigenic *Escherichia coli* (ETEC) seems to behave similarly, with deep
108 branches of the phylogeny obscured by frequent recombination and plasmid exchange, but more
109 recent branches experiencing mostly clonal descent, with tight linkage between virulence factors
110 and the genomic mainland [23]. These observations are consistent with an ancient, diverse
111 panmictic gene pool giving rise to clonal expansions, which can last for decades or centuries.
112

113 **The balance between recombination and selection**

114 Let us consider the evolutionary forces that determine clonality: natural selection and
115 recombination. The effect of recombination on clonality is straightforward: more recombination
116 means less clonality. The effect of natural selection is more complex, but is defined here simply
117 as a force which favors clonal expansions of adaptive mutants within an ecological niche. When
118 driven by ecological selection, clonal expansions are called selective sweeps, in which one clone
119 outcompetes all others, purging genetic diversity in the population [15,16].

120 Recombination and selection interact to determine the clonality of a population.
121 Recombination rates depend both on the ability of DNA to enter a cell and be incorporated into
122 the genome (the baseline rate) and the ability of that DNA to be retained by a balance of genetic
123 drift and natural selection (the realized rate). Some bacteria, such as *Helicobacter pylori,* have
124 realized recombination rates that are much higher than point mutation rates, exchanging at least

10% of their genome within a single four-year human infection [24]. Others, such as *Staphylococcus aureus* [25,26] and *Mycobacterium tuberculosis* [27-29] are decidedly more clonal. Recombination rates (both realized and baseline) vary widely across the genome. Of 10 pathogenic bacterial species studied, all had identifiable recombination 'hot' regions, although their length, genomic location and gene content varied [30]. Genes of different functions had different realized recombination rates, implying a role for natural selection on gene function in determining whether newly acquired genes are retained.

Modeling the recombination-selection balance

When rates of recombination are relatively low compared to rates of natural selection (s) on adaptive genes within niches, entire genomes will sweep to fixation before they can be shuffled by recombination. The $s >> r$ regime is well described in the Stable Ecotype Model [15], which predicts that most of the genome will follow a single, clonal phylogeny. Genome-wide sweeps thus increase clonality and can be considered a hallmark of clonal populations (**Table 1, Figure 1a**). In the $r >> s$ regime, individual genes (rather than entire genomes) will sweep to fixation (*i.e.* reach 100% frequency) in ecological niches to which they are adapted, without affecting genetic diversity elsewhere in the genome (**Figure 1b**). A recent model showed how these gene sweeps can occur at moderate (not unrealistically high) rates of recombination [18]. In this model, a microbial habitat is bombarded with genetically maladapted migrants, allowing gene sweeps to occur, although the adaptive allele never reaches 100% frequency due to the constant input of migrants. In another model, Takeuchi et al. [17] show that gene sweeps can occur when r is either very high or – counter-intuitively – when r is very low, but only when negative frequency-dependent selection (NFDS) reduces the rate of genome-wide selective sweeps. NFDS might be commonly imposed on bacteria and archaea by viral (phage) predation, providing a selective advantage to rare alleles of phage receptor genes [31].

Genome-wide and gene-specific sweeps in nature

To date, empirical evidence for gene-specific and genome-wide sweeps has come mostly from cross-sectional studies of a single population of genomes at a single point in time, with recombination and selection inferred backward in time [10-12,32]. Sequencing microbial genomes or metagenomes sampled over time – already a typical practice in genomic epidemiology (e.g. [26,33]) – promises to elucidate the rates of gene-specific and genome-wide sweeps in nature (**Figure 1**).

In a pioneering study, Bendall et al. [34] sampled a lake over 9 years and followed single-nucleotide polymorphism (SNP) and gene frequencies in 30 bacterial populations by metagenomic sequencing. They inferred that one of the populations (*Chlorobium*-111) had undergone a near-complete genome-wide sweep over the 9-year study, with most SNP diversity purged genome-wide (**Figure 1a**). In 6 other populations, they identified regions of the genome with unexpectedly low diversity compared to the genome-wide average, throughout the time-course. Either strong purifying selection on these regions or gene-specific selective sweeps (completed before the start of the time-course) could explain the low diversity. During the 9-year study, they also observed examples "where a few adjacent SNPs trended toward fixation while genome-wide diversity was maintained" (**Figure 1b**). They took this observation as consistent with gene-specific selective sweeps, but did not attempt to determine whether the sweeps were due to selection or drift. As a whole, the study showed that both genome-wide and gene-specific sweeps can occur in different microbial populations from the same environment. Whether microbial populations behaved differently due to differences in their ecology (*i.e.* regime of natural selection) or in their baseline recombination rates remains a question for future study. The fact that a genome-wide sweep was observed over a 9 year period suggests that such events might be relatively rapid but rare (only observed in 1 of 30 populations). Meanwhile, gene-sweeps might be more common historically (affecting 6 of 30 populations), but could take longer to proceed to completion.

177

178 **Is clonality a stable trait?**

179 As described in the *V. cholerae* example, some pathogenic bacterial populations can switch
180 between panmictic and clonal lifestyles. Therefore clonality can vary over time, but how much
181 and how often? To quantify the stability of clonality over time, I re-analyzed the lake time-course
182 of Bendall et al. [34]. Because estimates of selection and recombination rates were not readily
183 available for this dataset, I defined clonality based on the frequency of genome-wide selective
184 sweeps. Frequent genome-wide selective sweeps suggest $s >> r$, suggesting clonality. I identified
185 20 "old, diverse" populations as those with a high density of SNPs (>1500 SNPs/Mbp) at the
186 beginning of the time-course. These populations are likely "old" because they have gone a
187 relatively long time since the last genome-wide purge of genetic diversity and are relatively non-
188 clonal. They include the 6 populations inferred to have undergone gene-specific sweeps [34]. The
189 remaining 10 populations were defined as "young, low-diversity," having more recently
190 experienced a genome-wide purge of diversity. The "old, diverse" populations have a relatively
191 low ratio of nonsynonymous (N) to synonymous (S) SNPs, suggesting large effective population
192 sizes and ample time for purifying selection to remove (mostly deleterious) nonsynonymous
193 mutations (**Figure 2**). In contrast, the "young, low-diversity" populations are more likely to have
194 high N:S ratios, suggesting smaller effective population sizes and less time for purifying selection
195 to have acted.

196 With young (clonal) and old (less clonal) populations thus defined, I asked whether "old,
197 diverse" populations tended to maintain their diversity through the 9-year period of the study.
198 Bendall et al. defined two alternative population types: 1) those that maintained stable SNP
199 diversity over 9 years, and 2) those that experienced significant fluctuations in diversity due to
200 clonal expansions – defined when one, but not all timepoints are dominated by a single allele
201 (≥95% frequency) at >40% of SNP sites in the genome [34]. By definition, the 19 populations of
202 the first type did not experience genome-wide sweeps during the study, while the 11 populations

of the second type did experience genome-wide purges of diversity, which were transient in 10 cases and apparently permanent in 1 case (*Chlorobium*-111). Strikingly, 17 out of 20 "old, diverse" populations maintained their diversity over the 9-year study, compared to only 2 out of 10 "young, low-diversity" populations (Fisher test, Odds Ratio = 19.4, $P < 0.001$). This result suggests that populations with a history of genome-wide sweeps (and/or population bottlenecks) tend to experience subsequent genome-wide sweeps, and those that have maintained genetic diversity in the past tend to maintain their diversity into the future. In other words, clonality can be considered a relatively stable microbial trait.

History repeats itself

It appears that pathogens are more likely than free-living bacteria to undergo clonal expansions, due in part to their ecology and transmission dynamics [1,35]. Free-living aquatic bacteria, on the other hand, seem to be more likely to live in large, panmictic populations and behave like archipelagos [10-12,34]. If clonality is indeed a stable trait, this implies that history will repeat itself, and that the future behavior of microbial populations can be predicted with some confidence from their past behavior. Diverse populations tend to stay diverse. Clonal populations (that experience frequent genome-wide sweeps) tend to stay clonal. But history is not doomed to repeat itself forever. As we have seen, clonal expansions, such as pandemic *V. cholerae,* may originate when a panmictic gene pool (an archipelago) coalesces into a clonal continent, with virulence factors linked as peninsulas. Many such pathogenic clones have been documented, with life spans of decades to thousands of years [23,26,36,37]. Other pathogens, such as *Streptococcus pneumoniae*, may retain their panmictic population structure throughout an outbreak [38-40]. Why some pathogens are clonal and others are panmictic is an open question, but surely depends on the balance between recombination and selection, and on the time scales considered.

Acknowledgments. I am grateful to Yan Boucher and Salvador Almagro-Moreno for their thoughtful comments which improved the manuscript. I was supported by a Canada Research Chair.

Table 1. 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 [20,21]
Peninsula	Gene	Genome-wide	vertical (clonal)	the cholera toxin gene, acquired horizontally, then linked to a clonal *V. cholerae* genome [8,19]
Continent	Genome	Genome-wide	vertical (clonal)	clonal expansions of *S. aureus* [26] and *M. tuberculosis* [29,37]
Archipelago	Genome	Gene-specific	horizontal	hotspring cyanobacteria [10], ocean vibrios [11,12], pneumococcus [38,40]

Figures

Figure 1. Temporal dynamics of genome-wide and gene-specific selective sweeps inferred from metagenomic data. Genetic diversity can be measured by mapping metagenomic sequence reads to a reference genome, identifying SNPs, and calculating the allele frequencies at each SNP position in the genome over time. The lowest possible genetic diversity occurs when a single allele is present in 100% of metagenomic reads. Alternatively, diversity could be defined in terms of gene presence/absence, based on relative coverage of a gene in the reference genome by metagenomic reads. **(a)** In a hypothetical genome-wide selective sweep, all positions in the genome tend toward low diversity over time. **(b)** In a hypothetical gene-specific selective sweep, only one or a few positions in the genome tend toward low diversity, while the rest of the genome maintains high or intermediate diversity.

Figure 2. Past diversity predicts future diversity. Based on data from Table 2 of Bendall et al. [34], the genome-wide average nonsynonymous to synonymous (N:S) SNP ratio was plotted against the total SNP density (SNPs per megabasepair) for each of 30 bacterial populations. A pseudocount of 1 was added to both N and S counts. These 30 populations were divided into 20 "old, diverse" populations (>1500 SNPs/Mbp) and 10 "young, low-diversity" populations (<1500 SNPs/Mbp), highlighted in blue and yellow, respectively. Each point represents one of the 30 populations, colored in black if diversity was maintained over a 9-year metagenomic time-course, or in red if it was not. Seventeen out of 20 "old, diverse" populations maintained their diversity over the 9-year study, compared to only 2 out of 10 "young, low-diversity" populations (Fisher test, Odds Ratio = 19.4, $P < 0.001$). Consistent with previous observations that N:S depends on the evolutionary time available for purifying selection to act [41,42], N:S is negatively correlated with SNPs/Mbp, a proxy for evolutionary time or the time since the last genome-wide purge of genetic diversity in this dataset (Pearson's correlation of $\log_{10}$ transformed data, $r = -0.81$, $P = 5.6e\text{-}8$).

References

1. Smith JM, Smith NH, O'Rourke M, Spratt BG: **How clonal are bacteria?** *Proc. Natl. Acad. Sci. USA* 1993, **90**:4384–4388.

2. Guttman DS, Dykhuizen DE: **Clonal Divergence in Escherichia coli as a Result of Recombination, Not Mutation**. *Science* 1994, **266**:1380–1383.

3. Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H: **Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution.** *Mol Microbiol* 1997, **23**:1089–1097.

4. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF, Chisholm SW: **Genomic islands and the ecology and evolution of *Prochlorococcus***. *Science* 2006, **311**:1768–1770.

5. Coleman ML, Chisholm SW: **Ecosystem-specific selection pressures revealed through comparative population genomics**. *Proc. Natl. Acad. Sci. USA* 2010, **107**:18634–18639.

6. Shapiro BJ: **Signatures of natural selection and ecological differentiation in microbial genomes.** *Advances in experimental medicine and biology* 2014, **781**:339–359.

7. Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjögren K-G, Pedersen AG, Schubert M, Van Dam A, Kapel CMO, et al.: **Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago.** *Cell* 2015, **163**:571–582.

8. Waldor MKM, Mekalanos JJJ: **Lysogenic conversion by a filamentous phage encoding cholera toxin.** *Science* 1996, **272**:1910–1914.

9. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, Taviani E, Jeon Y-S, Kim DW, Lee J-H, et al.: **Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae***. *Proc. Natl. Acad. Sci. USA* 2009, **106**:15442–15447.

10. Rosen MJ, Davison M, Bhaya D, Fisher DS: **Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche.** *Science* 2015, **348**:1019–1023.

* Deep sequencing of 90 cyanobacterial marker genes reveals extensive recombination, with each genome composed of a random mix of alleles from the gene pool.

11. Cui Y, Yang X, Didelot X, Guo C, Li D, Yan Y, Zhang Y, Yuan Y, Yang H, Wang J, et al.: **Epidemic Clones, Oceanic Gene Pools, and Eco-LD in the Free Living Marine Pathogen *Vibrio parahaemolyticus***. *Mol Biol Evol* 2015, **32**:1396–1410.

* In-depth population genomic evidence of a panmictic ocean gene pool containing non-clonal ecological populations.

12. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabo G, Polz MF, Alm EJ: **Population Genomics of Early Events in the Ecological Differentiation of Bacteria**. *Science* 2012, **336**:48–51.

13. Vulić M, Dionisio F, Taddei F, Radman M: **Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria.** *Proc. Natl. Acad. Sci. USA* 1997, **94**:9763–9767.

14. Majewski JJ, Cohan FMF: **DNA sequence similarity requirements for interspecific recombination in *Bacillus*.** *Genetics* 1999, **153**:1525–1533.

15. Wiedenbeck J, Cohan FM: **Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches**. *FEMS Microbiology Reviews* 2011, **35**:957–976.

16. Shapiro BJ, Polz MF: **Ordering microbial diversity into ecologically and genetically cohesive units.** *Trends Microbiol* 2014, **22**:235–247.

* A comprehensive discussion of the roles of selection and recombination in structuring microbial diversity.

17. Takeuchi N, Cordero OX, Koonin EV, Kaneko K: **Gene-specific selective sweeps in bacteria and archaea caused by negative frequency-dependent selection**. *BMC Biology* 2015, **13**:20.

* Mathematical modeling shows how individual genes can sweep through populations with low recombination rates in the presence of negative frequency-dependent selection.

18. Niehus R, Mitri S, Fletcher AG, Foster KR: **Migration and horizontal gene transfer divide microbial genomes into multiple niches**. *Nature Communications* 2015, **6**:1–9.

19. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, Croucher NJ, Choi SY, Harris SR, Lebens M, et al.: **Evidence for several waves of global transmission in the seventh cholera pandemic**. *Nature* 2011, **477**:462–U111.

20. Orata FD, Kirchberger PC, Méheust R, Barlow EJ, Tarr CL, Boucher Y: **The Dynamics of Genetic Interactions between *Vibrio metoecus* and *Vibrio cholerae*, Two Close Relatives Co-Occurring in the Environment**. *Genome Biol Evol* 2015, **7**:2941–2954.

* A clear demonstration that genes in the integron are recombined between species just as often as within species, and that 24% of 'core' genes have crossed species boundaries.

21. Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Bapteste E, Lopez P, Tarr CL, Polz MF: **Local Mobile Gene Pools Rapidly Cross Species Boundaries To Create Endemicity within Global Vibrio cholerae Populations**. *mBio* 2011, **2**:e00335–10.

22. Boucher Y, Orata FD, Alam M: **The out-of-the-delta hypothesis: dense human populations in low-lying river deltas served as agents for the evolution of a deadly pathogen.** *Front Microbiol* 2015, **6**:1120.

23. Mentzer von A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, Rasko DA, Joffre E, Corander J, Pickard D, et al.: **Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution**. *Nature Genetics* 2014, **46**:1321–1326.

24. Cao Q, Didelot X, Wu Z, Li Z, He L, Li Y, Ni M, You Y, Lin X, Li Z, et al.: **Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis.** *Gut* 2015, **64**:554–561.

25. Vos M, Didelot X: **A comparison of homologous recombination rates in bacteria and archaea**. *The ISME journal* 2009, **3**:199–208.

26. Baines SL, Holt KE, Schultz MB, Seemann T, Howden BO, Jensen SO, van Hal SJ, Coombs GW, Firth N, Powell DR, et al.: **Convergent adaptation in the dominant global hospital clone ST239 of methicillin-resistant *Staphylococcus aureus*.** *mBio* 2015, **6**:e00080.

27. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, et al.: **High Functional Diversity in *Mycobacterium tuberculosis* Driven by Genetic Drift and Human Demography**. *Plos Biol* 2008, **6**:e311.

28. Namouchi A, Didelot X, Schock U, Gicquel B, Rocha EPC: **After the bottleneck: Genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection**. *Genome Research* 2012, **22**:721-734.

29. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant J, Parkhill J, Nejentsev S, et al.: **Evolution and transmission of drug-resistant tuberculosis in a Russian population**. *Nature Genetics* 2014, **46**:279–286.

30. Yahara K, Didelot X, Jolley KA, Kobayashi I, Maiden MCJ, Sheppard SK, Falush D: **The landscape of realized homologous recombination in pathogenic bacteria.** *Mol Biol Evol* 2015, doi:10.1093/molbev/msv237.

31. Cordero OX, Polz MF: **Explaining microbial genomic diversity in light of evolutionary ecology**. *Nature Reviews Microbiology* 2014, **12**:263-273.

32. Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, Krause DJ, Whitaker RJ: **Patterns of Gene Flow Define Species of Thermophilic Archaea**. *Plos Biol* 2012, **10**:e1001265.

33. Croucher NJ, Didelot X: **The application of genomics to tracing bacterial pathogen transmission**. *Current Opinion in Microbiology* 2015, **23**:62–67.

34. Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, et al.: **Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations**. *The ISME journal* 2016, **doi:10.1038/ismej.2015.241**

***** A pioneering application of time-course metagenomics to infer genome-wide and gene-specific selective sweeps in natural lake bacterial populations.

35. Achtman M, Wagner M: **Microbial diversity and the genetic nature of microbial species.** *Nature Reviews Microbiology* 2008, **6**:431–440.

36. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing Program,

Henderson DK, Palmore TN, Segre JA: **Tracking a Hospital Outbreak of Carbapenem-Resistant *Klebsiella pneumoniae* with Whole-Genome Sequencing**. *Science Translational Medicine* 2012, **4**:148ra116.

37. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, et al.: **Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans.** *Nature Genetics* 2013, **45**:1176–1182.

38. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M: **Population genomics of post-vaccine changes in pneumococcal epidemiology.** *Nature Genetics* 2013, **45**:656–663.

39. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, Pessia A, Aanensen DM, Mather AE, Page AJ, et al.: **Dense genomic sampling identifies highways of pneumococcal recombination.** *Nature Genetics* 2014, **46**:305–309.

40. Marttinen P, Croucher NJ, Gutmann MU, Corander J, Hanage WP: **Recombination produces coherent bacterial species clusters in both core and accessory genomes**. *Microbial Genomics* 2015, **DOI: 10.1099/mgen.0.000038**.

* The first simulation of bacterial evolution including mutation, allelic exchange (homologous recombination) and gene gain/loss (non-homologous recombination).

41. Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, Smith NH, Feil EJ: **Comparisons of dN/dS are time dependent for closely related bacterial genomes**. *J Theor Biol* 2006, **239**:226–235.

42. Kryazhimskiy S, Plotkin JB: **The Population Genetics of dN/dS**. *PLoS Genet* 2008, **4**:e1000304.

Figure 1

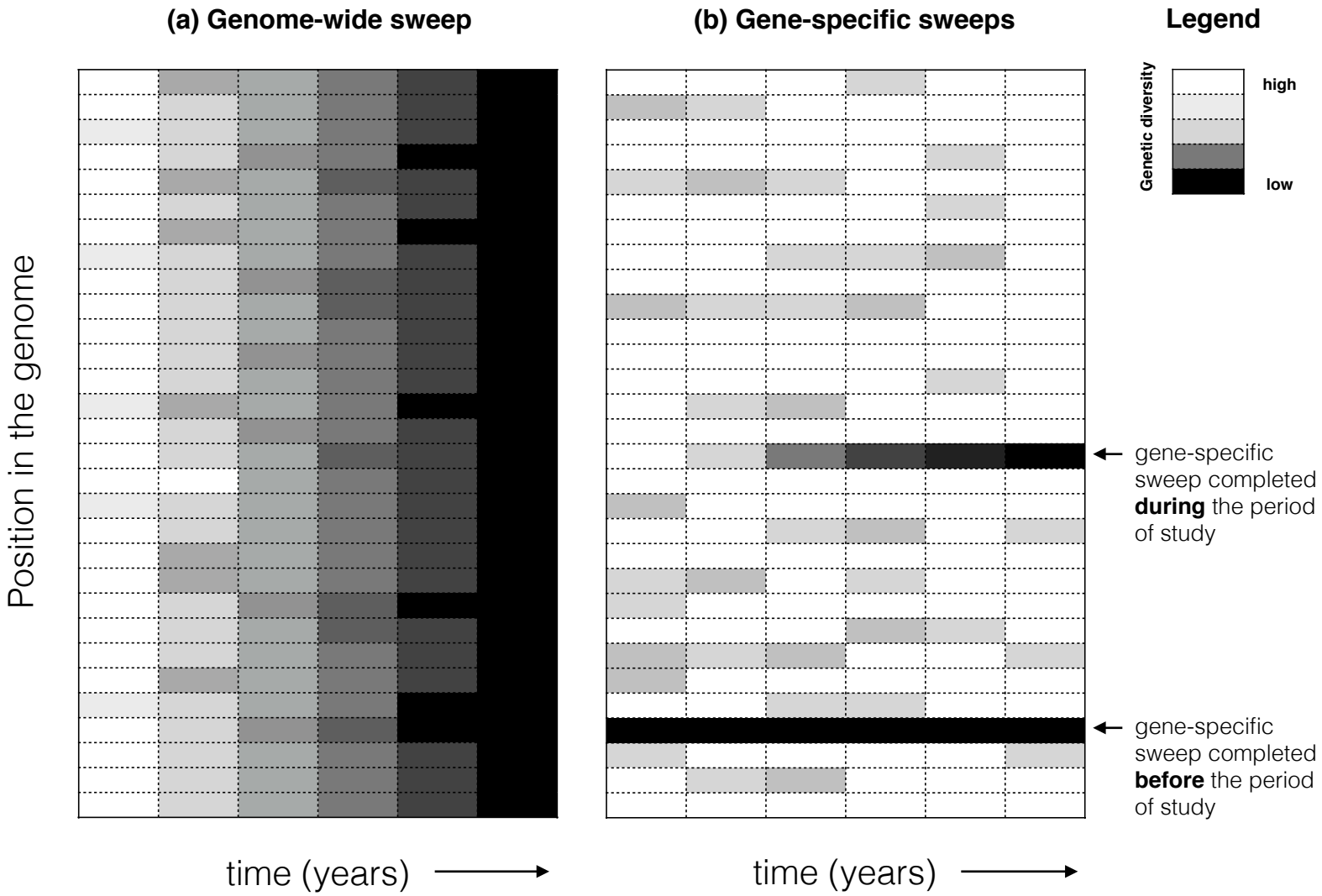

Figure 2