Massachusetts Institute of Technology

Detecting bacterial adaptation within individual microbiomes

Tami D. Lieberman[1,2,3]

[1]Department of Civil and Environmental Engineering, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Broad Institute, Cambridge, MA, USA
[3]Ragon Institute, Cambridge, MA, USA

TDL, 0000-0001-5430-3937

The human microbiome harbours a large capacity for within-person adaptive mutations. Commensal bacterial strains can stably colonize a person for decades, and billions of mutations are generated daily within each person's microbiome. Adaptive mutations emerging during health might be driven by selective forces that vary across individuals, vary within an individual, or are completely novel to the human population. Mutations emerging within individual microbiomes might impact the immune system, the metabolism of nutrients or drugs, and the stability of the community to perturbations. Despite this potential, relatively little attention has been paid to the possibility of adaptive evolution within complex human-associated microbiomes. This review discusses the promise of studying within-microbiome adaptation, the conceptual and technical limitations that may have contributed to an underappreciation of adaptive *de novo* mutations occurring within microbiomes to date, and methods for detecting recent adaptive evolution.

This article is part of a discussion meeting issue 'Genomic population structures of microbial pathogens'.

1. Introduction

While much attention has been paid to the ecology of microbiomes within individual people—which bacterial species reside within us during health [1,2], how these species interact with one another [3–5], and how they interact with the host [6,7]—relatively less attention has been paid to the evolution of single species and strains within these communities. Yet, there is a large potential for adaptive evolution within individual human microbiomes [8], enabled by population sizes greater than 10^{12} cells per person [9]. Mutations that confer antibiotic-resistance arise during infections and outcompete their ancestral genotype within days [10], and individual strains can persist and diversify within a person's microbiome for decades [11]. As such, individual strains have ample opportunity to explore potential adaptive mutations and change significantly during their residence within a host.

To date, studies of microbiome evolution have largely focused on relatively long timescales—how our microbiomes have changed as mammals [12], primates [13], or human populations diversified [14]. These studies have characterized how host evolution has selected for different bacterial communities, defined at the species or higher taxonomic level using 16S rRNA sequencing [15,16], but less attention has been paid to how those species and strains themselves might be changing. Even when studies focus on genomic changes within a species, these have mostly focused on timescales considerably longer than a single human lifetime, such as the absence of motility operons within a subspecies of *Eubacterium rectale* that is largely restricted to Europe [17].

Yet, tracking bacterial adaptation at short timescales, particularly within individual human microbiomes, can yield both practical and conceptual

insights. From an applied perspective, identifying genes and genomic loci undergoing adaptive evolution in real time provides insight into the selective pressure experienced by microbes *in vivo* [18,19] and therefore aids in the design of rational probiotic therapies. Identifying genetic variants critical to bacterial survival in human microbiomes can be difficult using *in vitro* and laboratory animal models. Bacterial gene expression and survival strategies are highly context-dependent, and it is not uncommon for small shifts in the environment, such as inclusion of a community member, to alter the degree to which a genetic mutation is advantageous or deleterious in a laboratory setting. As an extreme example, studies of how bacteria adapt to survive high levels of an antibiotic must be carefully designed to avoid selecting for mutations that enhance adhesion to the walls of a test tube [20]. Organs-on-chips [21] and mice with humanized microbiota [22], while powerful, do not mimic *in vivo* conditions well enough to avoid system-dependent selective pressures [23]. By contrast, rigorous genomic signatures of in-host adaptation can conclusively indicate the presence of selection [8,19]. The identification of such adaptive changes provides a starting point for hypothesis-generation regarding selective forces, follow-up investigations using experimental, computational and other approaches [24], and engineering of probiotic therapies. For example, a study of the commensal *Bacteroides fragilis* in healthy subjects revealed single nucleotide changes in importers of complex polysaccharides that provide strong selective advantages; these changes might reflect a metabolic challenge or, alternatively, pressure from phage that use these proteins as gateways into the cell [25]. Accordingly, the rational design of *Ba. fragilis* probiotics should include selection of strains with the ability to thrive in the presence of these selective pressures.

On a conceptual level, understanding the extent to which microbes evolve within individual microbiomes is critical to understanding microbe–disease association and building predictive models of microbiome assembly. Adaptive mutations emerging within microbiomes might impact how a bacterium interacts with the immune system, the metabolism of nutrients or drugs, or the stability of the community to perturbations. If adaptive mutations are common within microbiomes, understanding the impact of a particular species or strain on host health or community composition might require full-genome resolution. In support of this idea, a recent study identified that loss-of-function mutations in a particular gene of *Staphylococcus aureus* is associated with the inflammatory skin disease atopic dermatitis [26], even though no study has found strains or phylogroups of this species associated with disease. Moreover, adaptive mutations might change the way we think about microbiome stability and resistance to perturbation [5,27]. Adaptation within individual microbiomes might be driven by selective forces that are new to industrialized societies, vary across individuals, or vary within individual people. If person-specific selective forces are common, it is possible that the bacteria within each person's microbiome may have already adapted, at least in part, to unique selective forces in their community [28]; this early adaptation may explain the resistance of established microbiomes to invaders [2,11].

Lastly, tracking evolution at short timescales is critical for understanding the mechanisms and selection modes (e.g. purifying versus positive) responsible for genetic variation within the microbiome [29]. Rates of genetic change in bacteria, as well as signals for adaptive evolution, are highly dependent upon the timescale separating the compared genomes [30–33]—estimates are faster [32] and more biased towards amino acid changes [33] when comparing closely related organisms than more distant ones. Timescale dependence makes it difficult to predict the relative likelihoods of mutations versus recombination, adaptive versus deleterious alleles and ecological versus evolutionary responses to perturbations within individual microbiomes [31,34]. Tracking evolution within individual microbiomes provides an opportunity to avoid this timescale dependence and quantify such real-time evolution in natural settings. Understanding the forces that create genetic variation within microbiomes will inform the degree to which such variation should be considered when modelling microbiome ecology or predicting the impact of microbiome-based interventions.

It may seem surprising that, despite this potential, the first studies of within-person bacterial evolution in the microbiome at a genomic scale have only emerged within the past five years [8,35–38]. Within-person adaptation is probably underappreciated because of both theoretical misconceptions and technical limitations of the most popular approaches in the microbiome field. This review aims to reveal and dispel these misconceptions and describe one powerful roadmap for detecting within-person adaptation. Examples from both commensals and infectious disease will be used to build intuition, with a particular emphasis on chronic opportunistic infections of the cystic fibrosis (CF) lung, as the field of within-host infectious disease evolution is more established. No attempt will be made to comprehensively review all studies of within-host bacterial evolution; recent reviews focusing on bacterial pathogen evolution provide an excellent summary of within-host evolutionary trends, many of which also apply to commensal microbes [39–41]. Instead, this piece focuses on demystifying the processes and theoretical considerations for detecting within-microbiome adaptation, with the aim to accelerate work in this area.

2. Rates of evolutionary and ecological changes in microbiomes

Human microbiomes can be considered stable or dynamic, depending on the timescale and genomic resolution considered [2,11,42]. Each person has a unique set of microbial species across their total microbiome, though the functional capacity of microbiomes varies less from one person to another than the specific membership [1]. The species [2,42] and strain [11,43] composition of a person's gut microbiome is remarkably stable on the timescale of a year, though the relative abundances of these organisms can fluctuate substantially on daily timescales [42,44].

The stability of human microbiomes at the level of membership is thought to emerge in part from priority effects [45,46]—the ability of early colonizers to alter community formation and exclude later migrants [47]. Accordingly, perturbations that lower bacterial abundance, including antibiotic treatment, are associated with an influx of newly colonizing strains in the gut [5,48]. Adult siblings share few strains in common [11], suggesting that such disturbances may be relatively common, although the likelihood that these sibling microbiomes were seeded by different organisms in early life has not yet been resolved.

There are many levels of resolution at which one can define a strain or other subspecies grouping. In this review, the primary unit of organization is a 'strain', defined as a monophyletic grouping of very closely related genotypes inferred to have a single-cell common ancestor within approximately the past 100 years (figure 1c; this has also been termed a 'cloud' [49]). The mutational distance cut-off to define a strain depends on the molecular clock of the studied organism, but it is often on the order of 100 mutations across the whole genome, as molecular clock rates in commensal and pathogenic bacteria do not vary that much—from 0.5 to 8 mutations genome^{-1} yr^{-1}, with the slow-growing *Mycobacterium tuberculosis* on the low end and *S. aureus* on the high end [41]. Exceptions must be made for hypermutators that accumulate mutations at higher rates owing to defects in DNA repair [50], and regions of the genome undergoing recombination or horizontal gene transfer must be removed before defining molecular clock rates or strain boundaries.

Given this relatively slow molecular clock range, it may seem surprising that beneficial mutations can be supplied at high rates within individual microbiomes. However, mutation supply and the molecular clock are distinct concepts. Each bacterial cell residing within a microbiome can supply random mutations created during DNA replication each bacterial generation; only the surviving mutations along a single line of descent are measured in calculation of molecular clocks.

During the years and decades that a strain stably persists within a microbiome, its member organisms thus acquire *de novo* mutations that create new genotypes and 'substrains' (figure 1c). Strains with substantial substrain structure have been found in a variety of infections [49,51–54], gut commensals [8,37,38], and skin commensals [36,55,56]. The amount of diversity within a strain can be used to infer lower bounds on its duration of residence in a microbiome [8,52]. This approach can overestimate the duration of colonization if within-person populations are founded by closely related cells, rather than from a single-cell bottleneck. On the other hand, residence time can be overestimated following an adaptive sweep (figure 1d) or neutral population bottleneck. Notably, not all adaptive mutants that rise in frequency lead to population-wide sweeps; it is frequent for competition between adaptive alleles on different genomic backgrounds (figure 2), termed clonal interference [57] or soft sweeps [58,59], to lead to coexistence of substrains [8,52,60].

The relative infrequency of immigrating strains and competitive between-strain dynamics (figure 1f) observed to date in human microbiomes [11] suggests that opportunity remains for *de novo* mutations to contribute to competitive dynamics (e.g. response of community to a new phage). Adaptive substrain dynamics, including adaptive sweeps, diversification, and clonal interference, have been shown to be common for *Ba. fragilis* [8] and other species [11,37] in the human gut, as well as *S. aureus* on the skin of children with atopic dermatitis [61]. While detection of some of these dynamics is possible with metagenomics [11], approaches that yield high-quality single-cell genomes are required to build phylogenetic trees and get a full understanding of substrain structure. Current single-cell genomic techniques are too error prone and sparse in their coverage of the genome. Instead, culture-based approaches that use the fact that colonies are typically founded by single cells,

and that minimize growth in the laboratories, have proved successful for reconstructing within-person evolution [8,36,38,56,61,62]. Using such culture-based approaches, adaptive substrain dynamics have been uncovered even in the presence of multiple-strain colonization in humans and mice [37,60].

3. Asexual within-host populations evolve differently to classical eukaryotic populations

Fundamental differences between evolution of haploid organisms, with primarily asexual reproduction, and that of classic, sexually recombing diploid organisms, has led some to underestimate the potential of adaptive evolution during colonization of a single host. Well-established models of adaptation in classical eukaryotes [63] create a strong intuition: the potential of a population to adapt increases with the amount of nucleotide diversity present within that population (π). Adaptive potential can be predicted from nucleotide diversity in classic diploid populations, which undergo one recombination event per chromosome per generation, for two reasons. First, a higher nucleotide diversity enables rearrangement of existing alleles into more novel genotypes with the potential for enhanced fitness. Second, high nucleotide diversity in a recombining population is considered to reflect a large population size and therefore predict a high rate of emergence of new, potentially adaptive, *de novo* mutations in the population. It is no wonder, then, that extremely low levels of within-person nucleotide diversity [64] are sometimes treated as severe limitations on adaptive potential.

Yet, the shuffling of alleles via recombination is likely to be rare for most within-person bacterial populations. As bacteria can only undergo homologous recombination with closely related individuals, usually of the same species [65], the low number of coexisting strains of any given species within a given microbiome [11,37] limits the supply of alternative alleles for any given gene (horizontal gene transfer is a distinct process from recombination, which brings in new alleles of existing genes). Thus, even if a bacterial strain is naturally competent [66] or has the capacity for phage-mediated gene transfer [67,68], the potential for bringing in new alleles via recombination is low; most recombination events will occur within members of the same strain. This low level of recombination means that measures of nucleotide diversity and adaptive potential are decoupled; while a two-strain population has access to a larger genotypic space than a single-strain population, this increase is not proportional to the corresponding change in π.

This genome-wide linkage limits the number of unique genotypes that can be found within each host [69]. When a strongly adaptive mutation arises in the population, it is unlikely to recombine and be found on a genomic background other than on the one on which it arose; instead, it is likely to drag all mutations on its genomic background—including neutral ones—to high frequency in a process known as genetic draft [70,71] (figure 2). Genetic draft, also termed hitchhiking, is not unique to microbial systems [72], but its impact here is large owing to genome-wide linkage. In this way, an adaptive sweep can purge all existing diversity and produce a population founded from a recent single cell. Therefore, positive selection and genetic draft keep π low,

global within-species diversity

within-person, within-species diversity
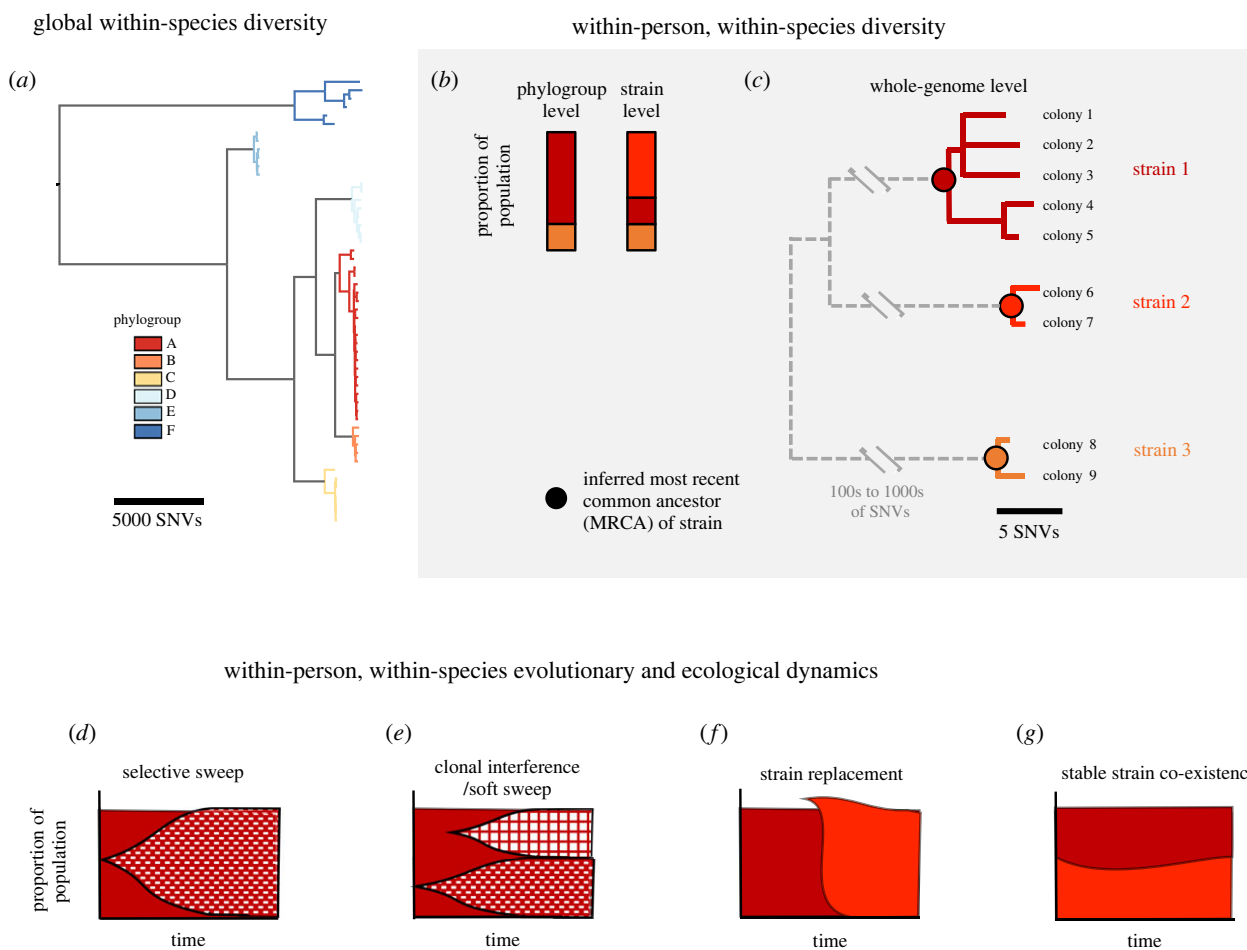
Figure 1. Subspecies classification and dynamics. (*a*) Within-species diversity of bacteria can be described as a phylogenetic tree and can be defined at multiple phylogenetic levels. Here, phylogroup is used to mean groupings of isolates separated by thousands of single nucleotide variants (SNVs) across the whole genome. (*b,c*) Within-person, within-species diversity can be described at multiple levels of resolution; the example has three strains of the studied species, two of which are from the same phylogroup and indicated in different shades of red. (*c*) Strains are defined here as groups of colonies separated by less than 100 SNVs across the genome; this cut-off is designed to group colonies that could have diversified on the host from a single cell; the exact cut-off used depends on the molecular clock rate of that species. (*d,e*) illustrate two types of subspecies evolutionary dynamics at the substrain level, while (*f,g*) illustrate ecological dynamics at the strain level.

despite years of colonization in a host, during which diversity could theoretically accumulate.

Instead, the greatest predictor of adaptive potential in within-host populations is likely to be the census population size—how many bacteria cells of a species are reproducing at a given time. This is because the supply of *de novo* mutations scales linearly with the number of cells in a population. For many bacterial species, the number of cells within an individual human host, and therefore the supply of mutations, is enormous. A minor species (0.1% abundance) in a person's gut can easily be comprised 10^{10} cells [9]. Given a conservative bacterial mutation rate of 10^{-10} mutations per cell division [73], such a population can explore every single point mutation across the genome each generation (probably days) [8].

4. Challenges in detecting recent within-host bacterial adaptation

Adaptation in asexual populations is also probably underestimated owing to technical challenges in distinguishing adaptive from neutral variants. A wide variety of statistical tests have been developed to identify genomic regions under recent pressure to change in classically sexual populations

[24,74]. For the most part, these tests compare diversity metrics within (such as π) or across populations (such as F_{ST}) to what would be expected under a neutral model with no adaptation. A widely acknowledged limitation of such tests for adaptation is that confounders, including spatial structure and changes in population size, can dramatically skew these metrics [24]. Therefore, it is standard practice to compare values of diversity metrics at a candidate locus to those obtained in the rest of the genome. This approach works because recombination in sexual organisms enables genomic regions to have independent diversity metrics. In primarily asexual organisms, however, loss of diversity at one site often can cause loss of diversity throughout the genome (figure 2). Without an internal control for these diversity metrics, it can be difficult to confirm a particular locus has recently undergone adaptive evolution.

The other major category of tools for identification of adaptive evolution relies on the fraction of mutations that are nonsynonymous (N), or amino acid changing, rather than synonymous (S), or amino acid preserving. Because of redundancy of the genetic code, only approximately 75% of mutations in coding regions are expected to change the amino acid code (the exact percentage depends on the codon usage and spectrum of mutations of a given organism). Therefore, adaptive evolution can be detected as a
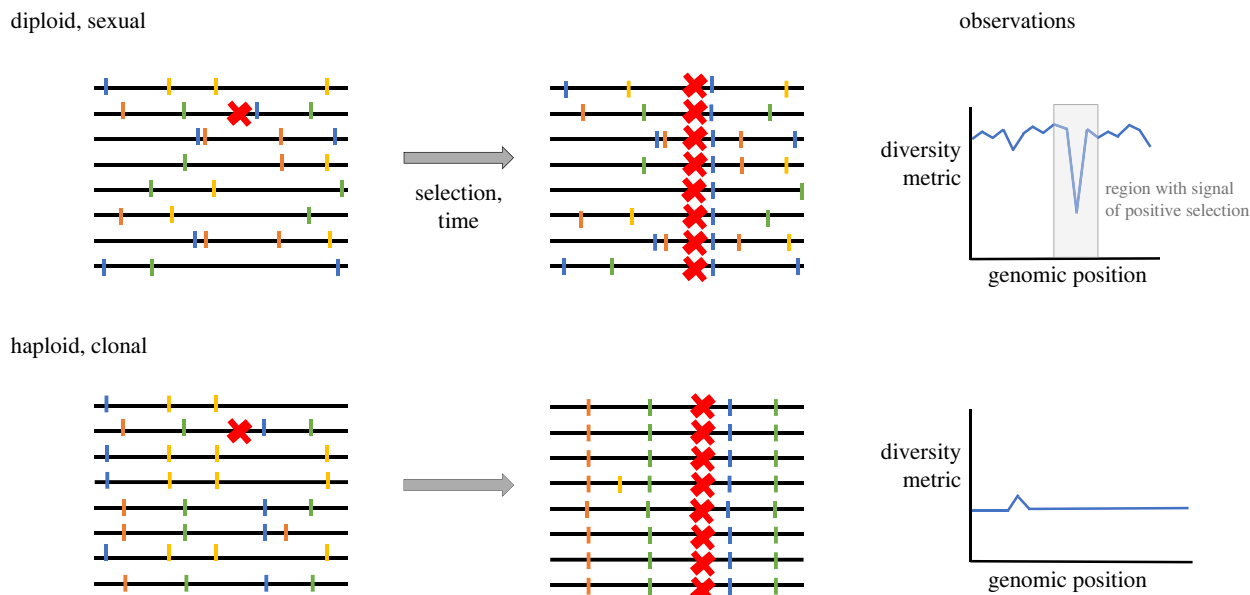
Figure 2. Recombination enables strong signatures for positive selection in recombining organisms but not in primarily asexual (clonal) populations. Recombination breaks up linkage between alleles in organisms that undergo frequent recombination. Red crosses indicate adaptive mutation. Other neutral alleles are represented by coloured vertical bars. When a selection event occurs, the adaptive mutation is driven to high frequency. In a highly recombining organism, this only drags very close alleles to high frequency, and thus a deviation from genome-wide metrics of diversity can be observed. However, in a primarily clonal population the population-wide sweep of an adaptive allele causes all diversity to be lost, at least until new mutations emerge. This process is known as genomic draft or hitchhiking.

statistically significant increase in the percentage of mutations that are non-synonymous relative to a calculated expectation (e.g. dN/dS) [75], to an expected value from between-species substitutions (e.g. McDonald-Kreitman) [76–78], or to observed values from other areas of the genome [79]. Critically, N versus S analyses must be done on a small segment of genomes, such as genes, protein domains, or even single nucleotides [78]. As most sites in a genome are under pressure to retain their amino acid sequence (purifying selection), signals of adaptive evolution can disappear when N versus S analyses are performed genome-wide [19].

However, a major challenge of N versus S approaches is the number of mutations needed to observe a statistically significant enrichment in the percent of mutations that are N (p_N). This difficulty arises because p_N is so high under neutral expectation. For example, to reject a null model of 75% N, at least 11 N mutations must be observed—under such a null model, 10 of 10 mutations being non-synonymous is not statistically significant to a $p = 0.05$ significance threshold (one-sided binomial test). This calculation only reflects a lower bound on the number of mutations, as it assumes no neutrally drafted or adaptive synonymous mutations in the gene of interest. Moreover, this simple calculation omits the necessary step of multiple-hypothesis correction for genome-wide scans. Therefore, using N/S ratios to detect genes undergoing adaptive evolution often requires dozens of observed mutations per gene, or more. When assessing within-person adaptation, it is rare to observe such a high number of mutations in a single gene. Typical datasets, even across cohorts of over a dozen individuals, include only hundreds or low thousands of *de novo* mutations across the genome [8,62,80]. Therefore, methods relying on p_N or dN/dS are better suited to the larger number of mutations generated on longer timescales or aggregating over sets of genes.

5. Parallel evolution as a powerful tool for detecting recent within-host adaptation

One approach with demonstrated success in detecting bacterial genes under adaptive evolution within individual hosts is scanning for parallel evolution (PE)—the emergence of similar mutations on distinct genomic backgrounds. Scans for PE start by counting the number of independent mutation events that occurred in a gene. This step requires phylogenetic analysis and is more complex than just identifying variable sites. Processes like recombination or repair of double-stranded DNA breaks can create multiple variant sites emerging from a single event [81,82]. On the other hand, a single variable site could emerge from two identical, parallel, mutations occurring in different genomic backgrounds [19,83]. Next, the number of mutations is compared to a neutral model in which mutations are randomly scattered throughout the genome [8,56]. While multiple-hypothesis correction remains an issue, approaches that use false-discovery rates (FDR), including simulations, have shown success in detecting meaningful signals of PE [8,52,84].

PE can only weakly support adaptive evolution—if considered in isolation. Signatures of PE could emerge if regions of the genome have intrinsically higher mutation rates. To confirm adaptive evolution, dN/dS or another secondary test can be applied to the set of genes with a signal of PE [8,19]. As tests such as dN/dS are difficult to interpret with small numbers of observed mutations, grouping all mutations within a set of genes under PE can provide sufficient signal. This two-step process is statistically valid because the number of mutations in a given gene and the proportion of those mutations which are N are independent. Thus, a set of genes with significant signals for adaptation using both approaches can confidently indicate the presence of adaptive evolution.

While the use of an FDR rather than strict hypothesis correction means that not all identified genes are truly under pressure to change *in vivo*, inspection of the function of genes can support the conclusion of adaptive evolution. For example, in a study of *B. fragilis* evolution in the gut of healthy people, 6 of 16 genes inferred to be under PE were all in the SusC/D family of complex polysaccharide importers (a significant enrichment relative to the per cent of genes with this annotation) [8], providing confidence that genes in this family are under pressure to change within individual healthy gut microbiomes. However, relying on gene function annotations currently remains challenging owing to the large number of unannotated or loosely annotated microbial genes; literature searches can often reveal functional roles not captured by ontologies [85]. An alternative signal that can provide confidence in the adaptive nature of observed mutations is the location of mutations within a given gene. For example, in a study of *Burkholderia dolosa* evolution during long-term infection of patients with CF, 12 of 17 mutations within the gene *fixL* fall within two small domains of this oxygen-dependent response regulator [19]. The clustering of these mutations, and the fact that none of the mutations created premature stop codons, provided confidence that *B. dolosa* is under pressure to tweak, but not break, this response regulator *in vivo* [86].

While searches for parallel adaptive evolution have traditionally considered single-nucleotide variants, larger genomic changes can also drive adaptive evolution, e.g. the gain of a mobile element or deletion of a gene [87,88]. However, the adaptive nature of a given mobile element change can be more difficult to identify from a genome-wide analysis because neutral expectations of the number of changes are difficult to model (as such changes are not clocklike) and there is no accepted independent test for selection that is analogous to dN/dS. Instead, focused investigation of candidate regions has proven successful in confirming the adaptive nature of horizontal gene transfer events [87,89].

6. The power of within-person parallel evolution

PE can be considered at genomic scales (e.g. gene versus pathway) and geographical scales (e.g. between people versus intra-person). This section discusses the data needed to detect within-host mutations and then explains the surprising success of within-individual PE relative to across-individual PE.

To identify variants that emerged during colonization, it is important to compare many genomes from each person and strain studied. These isolates can be collected over time, collected from the extant coexisting diversity within a host at a single time point, or both. Typical studies profiling cross-sectional diversity collect 10–100 colonies from each person at a given timepoint [8,26,36,62], with the ideal number to capture variation depending on the species and sampling site.

It is not sufficient to compare just one isolate each from many individuals. Thousands of mutations typically separate isolates from different individuals, most of which occurred in the distant past. Studies comparing bacterial genomes separated by various timescales have shown that adaptive signals strongly depend on the timescale studied [8,11,31,33,90]. One possible contributor to this timescale dependence is that proportionally more neutral mutations reach detection on longer timescales, via both random between-host transmission

bottlenecks and hitchhiking [71] on adaptive sweeps—which can overwhelm the few loci undergoing adaptive evolution. In addition, selective forces can vary over time, creating variation that obfuscates signatures of parallelism and adaptive evolution (figure 3) [91,92].

Once a list of variants occurring within each individual has been obtained, genomic scans for PE can be performed on the list of all mutations (across-individual PE) or within individual subjects (within-individual PE). The only difference is the null model used: either a single model in which all mutations found are scattered randomly across the genome or a separate null model for each subject. Surprisingly, higher enrichments for non-synonymous mutations have been found when performing analyses at the within-individual level as compared with the across-individual level. In the previously mentioned study of *B. fragilis* evolution in 12 healthy human gut microbiomes, 16 genes were identified to be mutated at least twice within one or more subjects, compared to just five expected under a neutral model (a density threshold was used to exclude excessively long genes) [8]. When the same set of mutations was considered for across-individual PE, eight genes fell below the threshold for PE while only two additional genes were implicated in PE. Interestingly, these two new genes found in the across-person analysis had a lower p_N than expected from a neutral model (though not significantly). Studies of other bacterial species have also found similar strength of within-individual PE scans [53,93], with less sensitivity at the across-host level [8,26,94].

Why should an enrichment of mutations be observed from within-person analyses but not when the same data is aggregated across subjects? One possibility may be that selective forces are person specific. In the study of *B. fragilis*, the SusC-family gene *ccfC* was mutated three times in subject 1 and no times in other subjects. This raises the possibility that *ccfC* was under pressure to change owing to the unique set of selective forces in subject 1—imposed by their genetics, their behaviour, their immune system, or their microbiome.

Alternatively, this apparent 'person-specificity' could arise from differences in genomic background of each person's bacterial strain. Genomic background can diversify the response to selective pressure in both obvious and subtle ways. Most simply, divergent strains of a species with different gene content might overcome the same challenge using different mechanisms [95]. More subtly, genetic interactions (epistasis) between small mutations can alter the course of evolution. An interesting example comes from a study of *Pseudomonas aeruginosa* evolution during long-term infection of 34 patients with CF [80]. Mutations in the transcriptional regulator gene *algU* were observed under PE across subjects in that study. Yet, these *algU* mutations only occurred on genomic backgrounds that had already obtained a mutation in the transcriptional regulator gene *mucA*. Even subtler yet, *de novo* mutations that have already swept a within-individual population prior to sampling might have already alleviated a selective pressure [26].

Because of the importance of ever-changing genomic backgrounds, it can be easiest to detect within-individual PE when multiple solutions to the same challenge are coexisting on the same genomic background—before any of these parallel adaptive alleles has 'won out'. Thus, coexisting adaptive alleles are easiest to observe via the sequencing of many
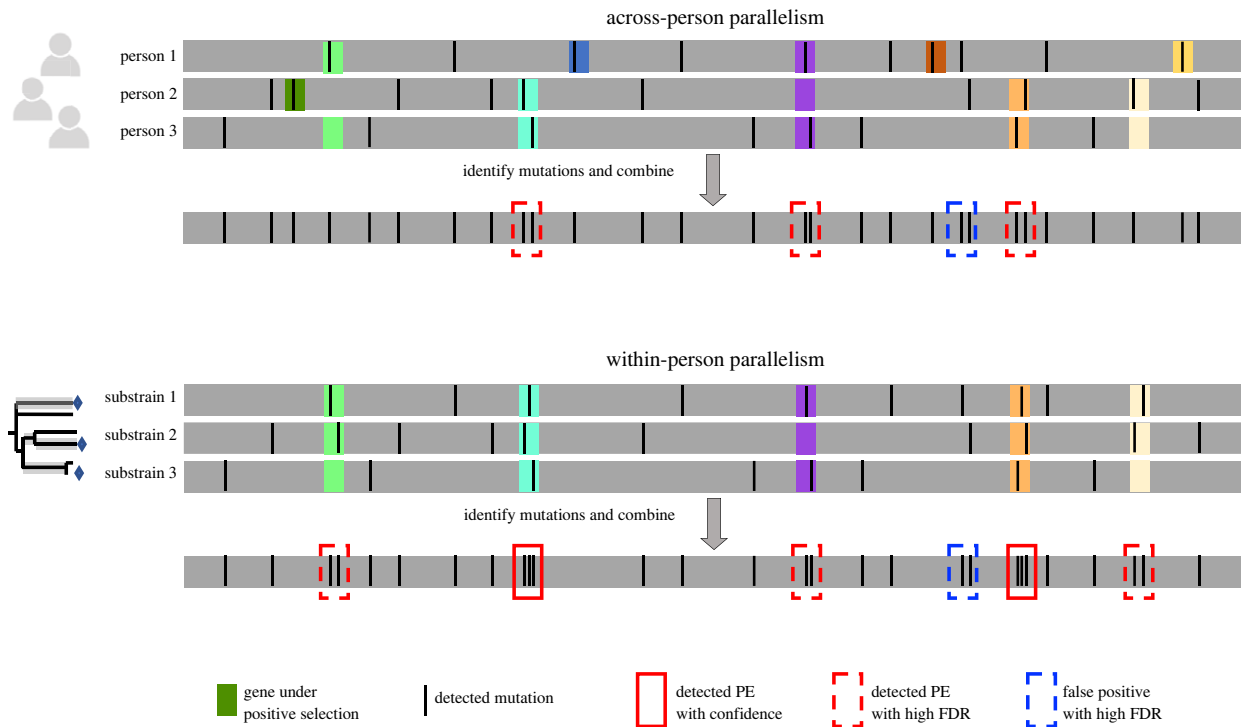
Figure 3. Across-person analyses lose statistical power in the presence of person-specific adaptation. Each line represents a bacterial genome from a single colony isolate. Shaded areas represent genes under positive selection shaded. In the top example, *de novo* mutations are identified separately in three subjects. As the genes under selection are not known ahead of time, a search for parallel evolution identifies only four genes mutated twice, and no genes mutated more times than that. One of these genes is a false positive (a gene not under directional selection). By contrast, when a single person is sampled more exhaustively such that more mutations which occurred within this subject are detected, two genes are confidently identified as under selection (three mutations each) owing to the uniformity of selective pressures across the sampled substrains. These examples illustrate how the presence of non-adaptive mutations creates statistical noise and a high bar for the confident detection of selection. Therefore, approaches which minimize the variation in selective forces between considered replicates have the most statistical power. Similar signal-to-noise issues can arise when comparing mutations at different stages of colonization or considering genes at the pathway level.

cross-sectional isolates from a single timepoint [52,96]. In this light, cross-sectional sampling can be more sensitive for detecting loci under adaptation than longitudinal sampling [8,52]. On the other hand, longitudinal studies can be used to estimate the selective advantage provided by a mutation [8,11], understand dynamics [8,10,59], and even observe reversions [19,80,92].

Identifying coexisting genotypes with alternative adaptive mutations is not possible for all species and environments. Not all bacterial populations enable competing adaptive mutations to persist for long enough to be detected, and bacterial diversity can be lost rapidly if population sizes are small [38] or a rare adaptive mutation is substantially more advantageous than other available mutations [97]. In these cases, detection of within-person PE may be challenging, and screens for across-individual PE (of within-person mutations) will be relatively more powerful. Across-individual PE is particularly useful when selective pressures are strong and uniform; this approach has been one of the primary signals for identifying mutations which confer antibiotic resistance *in vivo* [10,98,99].

7. What is the right genomic unit for detecting adaptation?

At the genomic level, scans for PE can use nucleotides, codons, genes, operons, or pathways as the operative unit. In theory, larger genomic units enable more potential for PE signals to reach genome-wide significance. On the other

hand, as genomic unit size increases, the chance also increases that the unit under consideration contains a mix of regions under directional selection (adaptive) and regions under purifying selection, thus erasing any signal [75]. An extreme example of this balancing effect comes from the previously mentioned study of *B. dolosa* evolution in infections of CF patients. In that study, genome-wide estimates of dN/dS were consistent with a neutral model even though a third of the intragenic mutations were concentrated in only 17 genes with highly elevated dN/dS. Signatures of purifying selection elsewhere in the genome neutralized the adaptive signal when averaged together [19]. Similarly, screens at the operon level only rarely detect cases of PE not detected at the gene level [52]. Yet, some new genes can be discovered at genomic units larger than the gene level, and it is relatively straightforward to test multiple genomic levels.

While, in theory, pathway-level screens for PE would remove some concerns regarding genome-specific solutions by categorizing mutations in broad functional groups, pathway-level annotations are not well-developed in microbial genomes. A large fraction of microbial genes remain hypothetical or unannotated, even in the most well-studied organisms. Even with pathway designations available, they are often rather coarse (e.g. transcriptional regulator instead of oxygen-dependent gene regulation). As a consequence, pathway sets in microbial genomes often contain very large sets of diverse genes, making it difficult to identify statistically significant signals.

For all these reasons, PE is easiest to detect using small genomic units and when positive selection is likely to be

uniform across studied genotypes. In practice, this is often the within-gene, within-person level—particularly when many genomes or colonies from the same time point can be sampled.

8. Conclusion

The future of the field of within-microbiome evolution is bright, with many questions left to address and many potential implications for understanding and manipulating microbiomes. In the coming years, studies across a wide variety of species inhabiting human microbiomes will reveal the degree to which adaptation drives within-person commensal evolution and dynamics, as well as the relevance of bacterial adaptation to health, interspecies interactions, and community stability.

On a theoretical level, studies of within-host selection will be critical to understanding evolution of microbiomes across longer timescales. It has long been appreciated that the ability of a bacterial mutant to survive and compete within a local environment (e.g. a single human microbiome) may not predict its fitness across larger ranges of environmental variation [100,101]. The degree to which within-person adaptation is person-specific [8] or short-sighted [92] remains to be determined. Regardless, much of commensal competition, and thus adaptation, operates at the level of within-host communities owing to limited between-host migrations [11]. As such, understanding within-host adaptive events will be critical to modelling and predicting the long-term success of bacterial strains [34].

Should *in vivo* adaptations be predictable within or across individuals, we may one day be able to predict mutations conferring immune escape, antibiotic resistance, or disease before they happen. Should adaptive mutations be predictable within an individual, but distinct across individuals, this will suggest a large need for personalization when engineering microbiomes.

Those species which have the most capacity to change *in vivo* may be drivers of microbiome dynamics, colonization resistance, or disease pathologies. Interestingly, not all studies of within-person microbiome evolution to date have observed strong signals of adaptive evolution. Studies of *S. aureus* evolution during asymptomatic nasal carriage in 13 subjects [62], *Escherichia coli* in the gut of one healthy subject [38], and *Cutibacterium acnes* on normal facial skin of 16 subjects [56] have revealed only signals of purifying selection. While these organisms have relatively lower population sizes than *Bacteroides* [8] and pathogens of the CF lung [19,80], they are still thought to have population sizes greater than the bacterial per-base-pair mutation rate. For *C. acnes*, it has been speculated that adaptive potential is limited by the highly structured environment in which it resides, including genotype-agnostic single-cell bottlenecks within individual pores [56]. However, the roles of predation by phage, environmental variability and molecular mechanisms of non-genetic adaptation may also play a role in the balance of adaptive and neutral evolution. Understanding the factors that determine adaptive potential will focus efforts seeking to link in-person mutations to health and disease.

References

1. Huttenhower C et al. 2012 Structure, function and diversity of the healthy human microbiome. Nature 486, 207–214. (doi:10.1038/nature11234)

2. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, Erdman SE, Alm EJ. 2014 Host lifestyle affects human microbiota on daily timescales. Genome Biol. 15, R89. (doi:10.1186/gb-2014-15-7-r89)

3. Coyte KZ, Rao C, Rakoff-Nahoum S, Foster KR. 2021 Ecological rules for the assembly of microbiome communities. PLoS Biol. 19, e3001116. (doi:10.1371/journal.pbio.3001116)

4. Friedman J, Alm EJ. 2012 Inferring correlation networks from genomic survey data. PLoS Comput. Biol. 8, e1002687. (doi:10.1371/journal.pcbi.1002687)

5. Koskella B, Hall LJ, Metcalf CJE. 2017 The microbiome beyond the horizon of ecological and evolutionary theory. Nat. Ecol. Evol. 1, 1606–1615. (doi:10.1038/s41559-017-0340-2)

6. Silpe JE, Balskus EP. 2021 Deciphering human microbiota–host chemical interactions. ACS Central Sci. 7, 20–29. (doi:10.1021/acscentsci.0c01030)

7. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ. 2013 Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. Sci. New York NY 341, 295–298. (doi:10.1126/science.1235872)

8. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, Xavier RJ, Alm EJ. 2019 Adaptive evolution within gut microbiomes of healthy people. Cell Host Microbe 25, 656–667. (doi:10.1016/j.chom.2019.03.007)

9. Sender R, Fuchs S, Milo R. 2016 Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. 14, e1002533. (doi:10.1371/journal.pbio.1002533)

10. Snitkin ES, Zelazny AM, Gupta J, Program NCS, Palmore TN, Murray PR, Segre JA. 2013 Genomic insights into the fate of colistin resistance and Acinetobacter baumannii during patient treatment. Genome Res. 23, 1155–1162. (doi:10.1101/gr.154328.112)

11. Garud NR, Good BH, Hallatschek O, Pollard KS. 2019 Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol. 17, e3000102. (doi:10.1371/journal.pbio.3000102)

12. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, Alm EJ. 2017 Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nat. Commun. 8, 14319. (doi:10.1038/ncomms14319)

13. Moeller AH et al. 2016 Cospeciation of gut microbiota with hominids. Science 353, 380–382. (doi:10.1126/science.aaf3951)

14. Groussin M et al. 2021 Elevated rates of horizontal gene transfer in the industrialized human microbiome. Cell 184, 2053–2067. (doi:10.1016/j.cell.2021.02.052)

15. Groussin M, Mazel F, Alm EJ. 2020 Co-evolution and co-speciation of host-gut bacteria systems. Cell Host Microbe 28, 12–22. (doi:10.1016/j.chom.2020.06.013)

16. Kolodny O, Callahan BJ, Douglas AE. 2020 The role of the microbiome in host evolution. Phil. Trans. R. Soc. B 375, 20190588. (doi:10.1098/rstb.2019.0588)

17. Karcher N et al. 2020 Analysis of 1321 Eubacterium rectale genomes from metagenomes uncovers

complex phylogeographic population structure and subspecies functional adaptations. *Genome Biol.* **21**, 138. (doi:10.1186/s13059-020-02042-y)

18. Cooper VS. 2018 Experimental evolution as a high-throughput screen for genetic adaptations. *Msphere* **3**, e00121-18. (doi:10.1128/msphere.00121-18)

19. Lieberman TD *et al.* 2011 Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat. Genet.* **43**, 1275–1280. (doi:10.1038/ng.997)

20. Toprak E, Veres A, Michel J-B, Chait R, Hartl DL, Kishony R. 2012 Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44**, 101–105. (doi:10.1038/ng.1034)

21. Jalili-Firoozinezhad S *et al.* 2019 A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. *Nat. Biomed. Eng.* **3**, 520–531. (doi:10.1038/s41551-019-0397-0)

22. Cheng AG *et al.* 2021 *In vivo* augmentation of a complex gut bacterial community. *Biorxiv*, 2021.06.15.448620. (doi:10.1101/2021.06.15.448620)

23. Vasquez KS *et al.* 2020 Quantifying the interplay between rapid bacterial evolution within the mouse intestine and transmission between hosts. *Biorxiv*, 2020.12.04.412072. (doi:10.1101/2020.12.04.412072)

24. Vitti JJ, Grossman SR, Sabeti PC. 2013 Detecting natural selection in genomic data. *Annu. Rev. Genet.* **47**, 97–120. (doi:10.1146/annurev-genet-111212-133526)

25. Shkoporov AN, Khokhlova EV, Stephens N, Hueston C, Seymour S, Hryckowian AJ, Scholz D, Ross RP, Hill C. 2021 Long-term persistence of crAss-like phage crAss001 is associated with phase variation in *Bacteroides intestinalis*. *BMC Biol.* **19**, 163. (doi:10.1186/s12915-021-01084-3)

26. Key FM *et al.* 2021 On-person adaptive evolution of *Staphylococcus aureus* during atopic dermatitis increases disease severity. *Biorxiv*, 2021.03.24.436824. (doi:10.1101/2021.03.24.436824)

27. Ferreiro A, Crook N, Gasparrini AJ, Dantas G. 2018 Multiscale evolutionary dynamics of host-associated microbiomes. *Cell* **172**, 1216–1227. (doi:10.1016/j.cell.2018.02.015)

28. Leibold MA, Urban MC, Meester LD, Klausmeier CA, Vanoverbeke J. 2019 Regional neutrality evolves through local adaptive niche evolution. *Proc. Natl Acad. Sci. USA* **116**, 201808615. (doi:10.1073/pnas.1808615116)

29. Feil EJ. 2004 Small change: keeping pace with microevolution. *Nat. Rev. Microbiol.* **2**, 483–495. (doi:10.1038/nrmicro904)

30. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. 2014 Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat. Commun.* **5**, 5471. (doi:10.1038/ncomms6471)

31. Shoemaker WR, Chen D, Garud NR. 2021 Comparative population genetics in the human gut microbiome. *Genome Biol. Evol.* **116**, evab116. (doi:10.1093/gbe/evab116)

32. Duchêne S, Holt KE, Weill F-X, Hello SL, Hawkey J, Edwards DJ, Fourment M, Holmes EC. 2016 Genome-scale rates of evolutionary change in bacteria. *Microb. Genom.* **2**, e000094. (doi:10.1099/mgen.0.000094)

33. Rocha EPC, Smith JM, Hurst LD, Holden MTG, Cooper JE, Smith NH, Feil EJ. 2006 Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J. Theor. Biol.* **239**, 226–235. (doi:10.1016/j.jtbi.2005.08.037)

34. Graves CJ, Weinreich DM. 2016 Variability in fitness effects can preclude selection of the fittest. *Annu. Rev. Ecol. Evol. Syst.* **48**, 1–19. (doi:10.1146/annurev-ecolsys-110316-022722)

35. Conwill A, Kuan AC, Damerla R, Poret AJ, Baker JS, Tripp AD, Alm EJ, Lieberman TD. 2021 Anatomy promotes neutral coexistence of strains in the human skin microbiome. *Biorxiv*, 2021.05.12.443817. (doi:10.1101/2021.05.12.443817)

36. Zhou W, Spoto M, Hardy R, Guan C, Fleming E, Larson PJ, Brown JS, Oh J. 2020 Host-specific evolutionary and transmission dynamics shape the functional diversification of *Staphylococcus epidermidis* in human skin. *Cell* **180**, 454–470. (doi:10.1016/j.cell.2020.01.006)

37. Poyet M *et al.* 2019 A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat. Med.* **25**, 1442–1452. (doi:10.1038/s41591-019-0559-3)

38. Ghalayini M, Launay A, Bridier-Nahmias A, Clermont O, Denamur E, Lescat M, Tenaillon O. 2018 Evolution of a dominant natural isolate of *Escherichia coli* in the human gut over the course of a year suggests a neutral evolution with reduced effective population size. *Appl. Environ. Microb.* **84**, e02377-17. (doi:10.1128/aem.02377-17)

39. Gatt YE, Margalit H. 2020 Common adaptive strategies underlie within-host evolution of bacterial pathogens. *Mol. Biol. Evol.* **38**, msaa278. (doi:10.1093/molbev/msaa278)

40. Sheppard SK, Guttman DS, Fitzgerald JR. 2018 Population genomics of bacterial host adaptation. *Nat. Rev. Genet.* **19**, 549–565. (doi:10.1038/s41576-018-0032-z)

41. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. 2016 Within-host evolution of bacterial pathogens. *Nat. Rev. Microbiol.* **14**, 150–162. (doi:10.1038/nrmicro.2015.13)

42. Gibbons SM, Kearney SM, Smillie CS, Alm EJ. 2017 Two dynamic regimes in the human gut microbiome. *PLoS Comput. Biol.* **13**, e1005364. (doi:10.1371/journal.pcbi.1005364)

43. Faith JJ *et al.* 2013 The long-term stability of the human gut microbiota. *Science* **341**, 1237439. (doi:10.1126/science.1237439)

44. Roodgar M *et al.* 2021 Longitudinal linked-read sequencing reveals ecological and evolutionary responses of a human gut microbiome during antibiotic treatment. *Genome Res.* **31**, 1433–1446. (doi:10.1101/gr.265058.120)

45. Martínez I *et al.* 2018 Experimental evaluation of the importance of colonization history in early-life gut microbiota assembly. *Elife* **7**, e36521. (doi:10.7554/elife.36521)

46. Sprockett D, Fukami T, Relman DA. 2018 Role of priority effects in the early-life assembly of the gut microbiota. *Nat. Rev. Gastroenterol.* **15**, 197–205. (doi:10.1038/nrgastro.2017.173)

47. Fukami T. 2012 Historical contingency in community assembly: integrating niches, species pools, and priority Effects. *Annu. Rev. Ecol. Evol. Syst.* **46**, 1–23. (doi:10.1146/annurev-ecolsys-110411-160340)

48. Smillie CS *et al.* 2018 Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal microbiota transplantation. *Cell Host Microbe* **23**, 229–240. (doi:10.1016/j.chom.2018.01.003)

49. Harris SR *et al.* 2013 Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect. Dis.* **13**, 130–136. (doi:10.1016/s1473-3099(12)70268-2)

50. Oliver A, Mena A. 2010 Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. *Clin. Microbiol. Infect.* **16**, 798–808. (doi:10.1111/j.1469-0691.2010.03250.x)

51. Lieberman TD, Wilson D, Misra R, Xiong LL, Moodley P, Cohen T, Kishony R. 2016 Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated *Mycobacterium tuberculosis*. *Nat. Med.* **22**, 1470–1474. (doi:10.1038/nm.4205)

52. Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, Kishony R. 2013 Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat. Genet.* **46**, 82–87. (doi:10.1038/ng.2848)

53. Chung H, Lieberman TD, Vargas SO, Flett KB, McAdam AJ, Priebe GP, Kishony R. 2017 Global and local selection acting on the pathogen *Stenotrophomonas maltophilia* in the human lung. *Nat. Commun.* **8**, 14078. (doi:10.1038/ncomms14078)

54. Caballero JD *et al.* 2015 Selective sweeps and parallel pathoadaptation drive *Pseudomonas aeruginosa* evolution in the cystic fibrosis lung. *Mbio* **6**, e00981-15. (doi:10.1128/mbio.00981-15)

55. Coll F *et al.* 2020 Definition of a genetic relatedness cutoff to exclude recent transmission of meticillin-resistant *Staphylococcus aureus*: a genomic epidemiology analysis. *Lancet Microbe* **1**, e328–e335. (doi:10.1016/s2666-5247(20)30149-x)

56. Conwill A, Kuan AC, Damerla R, Poret AJ, Baker JS, Tripp AD, Alm EJ, Lieberman TD. 2022 Anatomy promotes neutral coexistence of strains in the human skin microbiome. *Cell Host Microbe* **30**, 171–182. (doi:10.1016/j.chom.2021.12.007)

57. Gerrish PJ, Lenski RE. 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* **127**, 102–103. (doi:10.1023/a:1017067816551)

58. Hermisson J, Pennings PS. 2005 Soft sweeps. *Genetics* **169**, 2335–2352. (doi:10.1534/genetics.104.036947)

59. Barroso-Batista J, Sousa A, Lourenço M, Bergman M-L, Sobral D, Demengeot J, Xavier KB, Gordo I. 2013 The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. *PLoS Genet.* **10**, e1004182. (doi:10.1371/journal.pgen. 1004182)

60. Frazão N, Konrad A, Güleresi D, Lässig M, Gordo I. 2022 Two modes of evolution shape bacterial strain diversity in the gut for thousands of generations. *Biorxiv*, 2022.01.11.475860. (doi:10.1101/2022.01. 11.475860)

61. Key FM et al. 2021 On-person adaptive evolution of *Staphylococcus aureus* during atopic dermatitis increases disease severity. *Biorxiv*.

62. Golubchik T et al. 2013 Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage. *PLoS ONE* **8**, e61319. (doi:10.1371/journal. pone.0061319)

63. Wills C. 1998 Population genetics: a concise guide. John H. Gillespie. *Q. Rev. Biol.* **73**, 505–506. (doi:10. 1086/420456)

64. Schloissnig S et al. 2013 Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45–50. (doi:10.1038/nature11711)

65. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Carnevali PBM, Banfield JF. 2020 Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. *mSystems* **5**, e00731-19. (doi:10.1128/msystems.00731-19)

66. Vos M. 2009 Why do bacteria engage in homologous recombination? *Trends Microbiol.* **17**, 226–232. (doi:10.1016/j.tim.2009.03.001)

67. Marshall CW, Gloag ES, Lim C, Wozniak DJ, Cooper VS. 2021 Rampant prophage movement among transient competitors drives rapid adaptation during infection. *Sci Adv* **7**, eabh1489. (doi:10.1126/sciadv. abh1489)

68. Humphrey S, Fillol-Salom A, Quiles-Puchalt N, Ibarra-Chávez R, Haag AF, Chen J, Penadés JR. 2021 Bacterial chromosomal mobility via lateral transduction exceeds that of classical mobile genetic elements. *Nat. Commun.* **12**, 6509. (doi:10.1038/ s41467-021-26004-5)

69. Rocha EPC. 2018 Neutral theory, microbial practice: challenges in bacterial population genetics. *Mol. Biol. Evol.* **35**, 1338–1347. (doi:10.1093/molbev/ msy078)

70. Gillespie JH. 2000 Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* **155**, 909–919. (doi:10.1093/genetics/155.2.909)

71. Neher RA. 2013 Genetic draft, selective interference, and population genetics of rapid adaptation. *Ecol. Evol. Syst.* **44**, 195–215. (doi:10.1146/annurev-ecolsys-110512-135920)

72. Neher RA, Shraiman BI. 2011 Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics* **188**, 975–996. (doi:10.1534/ genetics.111.128876)

73. Lee H, Popodi E, Tang H, Foster PL. 2012 Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **109**, E2774–E2783. (doi:10.1073/pnas. 1210309109)

74. Field Y et al. 2016 Detection of human adaptation during the past 2000 years. *Science* **354**, 760–764. (doi:10.1126/science.aag0776)

75. Yang Z, Bielawski JP. 2000 Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503. (doi:10.1016/s0169-5347(00)01994-7)

76. Messer PW, Petrov DA. 2013 Frequent adaptation and the McDonald–Kreitman test. *Proc. Natl Acad. Sci. USA* **110**, 8615–8620. (doi:10.1073/pnas. 1220835110)

77. McDonald JH, Kreitman M. 1991 Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652–654. (doi:10.1038/351652a0)

78. Barber MF, Elde NC. 2014 Escape from bacterial iron piracy through rapid evolution of transferrin. *Science* **346**, 1362–1366. (doi:10.1126/science.1259329)

79. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. 2004 Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120. (doi:10.1086/ 421051)

80. Marvig RL, Sommer LM, Molin S, Johansen HK. 2015 Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat. Genet.* **47**, 57–64. (doi:10.1038/ng. 3148)

81. Didelot X, Wilson DJ. 2015 ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041. (doi:10. 1371/journal.pcbi.1004041)

82. Croucher NJ et al. 2011 Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434. (doi:10.1126/science. 1198545)

83. Lenski RE. 2011 Chance and necessity in the evolution of a bacterial pathogen. *Nat. Genet.* **43**, 1174–1176. (doi:10.1038/ng.1011)

84. Giulieri SG et al. 2022 Niche-specific genome degradation and convergent evolution shaping *Staphylococcus aureus* adaptation during severe infections. *Biorxiv*, 2022.02.11.480068. (doi:10.1101/ 2022.02.11.480068)

85. Price MN, Arkin AP. 2017 PaperBLAST: text mining papers for information about homologs. *mSystems* **2**, e00039-17. (doi:10.1128/msystems.00039-17)

86. Schaefers MM, Liao TL, Boisvert NM, Roux D, Yoder-Himes D, Priebe GP. 2017 An oxygen-sensing two-component system in the *Burkholderia cepacia* complex regulates biofilm, intracellular invasion, and pathogenicity. *PLoS Pathog.* **13**, e1006116. (doi:10.1371/journal.ppat.1006116)

87. Arnold BJ, Huang I-T, Hanage WP. 2021 Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* **20**, 206–218. (doi:10.1038/ s41579-021-00650-4)

88. Pearson T et al. 2020 Pathogen to commensal? Longitudinal within-host population dynamics, evolution, and adaptation during a chronic >16-year *Burkholderia pseudomallei* infection. *PLoS Pathog.* **16**, e1008298. (doi:10.1371/journal.ppat. 1008298)

89. Ross BD et al. 2019 Human gut bacteria contain acquired interbacterial defense systems. *Nature* **575**, 224–228. (doi:10.1038/s41586-019-1708-z)

90. Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Coper A. 2011 Time-dependent rates of molecular evolution. *Mol. Ecol.* **20**, 3087–3101. (doi:10.1111/j.1365-294x.2011. 05178.x)

91. Culyba MJ, Tyne DV. 2021 Bacterial evolution during human infection: adapt and live or adapt and die. *PLoS Pathog.* **17**, e1009872. (doi:10.1371/journal. ppat.1009872)

92. Sousa A, Ramiro RS, Barroso-Batista J, Güleresi D, Lourenço M, Gordo I. 2017 Recurrent reverse evolution maintains polymorphism after strong bottlenecks in commensal gut bacteria. *Mol. Biol. Evol.* **34**, 2879–2892. (doi:10.1093/molbev/msx221)

93. Silva IN et al. 2016 Long-term evolution of *Burkholderia multivorans* during a chronic cystic fibrosis infection reveals shifting forces of selection. *mSystems* **1**, e00029-16. (doi:10.1128/msystems. 00029-16)

94. Zdziarski J et al. 2010 Host imprints on bacterial genomes—rapid, divergent evolution in individual patients. *PLoS Pathog.* **6**, e1001078. (doi:10.1371/ journal.ppat.1001078)

95. Vogwill T, Kojadinovic M, Furió V, MacLean RC. 2014 Testing the role of genetic background in parallel evolution using the comparative experimental evolution of antibiotic resistance. *Mol. Biol. Evol.* **31**, 3314–3323. (doi:10.1093/molbev/msu262)

96. Feliziani S, Marvig RL, Luján AM, Moyano AJ, Rienzo JAD, Johansen HK, Molin S, Smania AM. 2014 Coexistence and within-host evolution of diversified lineages of hypermutable *Pseudomonas aeruginosa* in long-term cystic fibrosis infections. *PLoS Genet.* **10**, e1004651. (doi:10.1371/journal.pgen.1004651)

97. Fogle CA, Nagle JL, Desai MM. 2008 Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics* **180**, 2163–2173. (doi:10.1534/genetics.108.090019)

98. Farhat MR et al. 2013 Genomic analysis identifies targets of convergent positive selection in drug resistant *Mycobacterium tuberculosis*. *Nat. Genet.* **45**, 1183–1189. (doi:10.1038/ng.2747)

99. Baker S, Thomson N, Weill F-X, Holt KE. 2017 Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Sci New York NY* **360**, 733–738. (doi:10.1126/science. aar3777)

100. Kussell E, Leibler S. 2005 Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078. (doi:10. 1126/science.1114383)

101. Eshel I. 1973 Clone selection and the evolution of modifying features. *Theor. Popul. Biol.* **4**, 196–208. (doi:10.1016/0040-5809(73)90029-4)