Trends in Microbiology

CellPress

Review

Genome Surfing As Driver of Microbial Genomic Diversity

Mallory J. Choudoir,[1] Kevin Panke-Buisse,[1] Cheryl P. Andam,[2] and Daniel H. Buckley[1,*]

Historical changes in population size, such as those caused by demographic range expansions, can produce nonadaptive changes in genomic diversity through mechanisms such as gene surfing. We propose that demographic range expansion of a microbial population capable of horizontal gene exchange can result in genome surfing, a mechanism that can cause widespread increase in the pan-genome frequency of genes acquired by horizontal gene exchange. We explain that patterns of genetic diversity within *Streptomyces* are consistent with genome surfing, and we describe several predictions for testing this hypothesis both in *Streptomyces* and in other microorganisms.

Effects of Horizontal Gene Transfer (HGT) on Genomic Diversity

Microbial species and their populations exhibit remarkable genomic diversity. While mutation and recombination promote genetic variation in all forms of life, the genomic diversity of Bacteria and Archaea is enhanced dramatically by their proclivity for HGT. Genome sequencing of *Escherichia coli* provided one of the earliest demonstrations of the vast genomic diversity encompassed by microbial species [1–3]. Diverse strains of *Escherichia coli* share 63% of their genome on average with all other members of the species, while the accessory gene repertoire, often found in only one or few strains, encompasses greater than 5× the core gene content [4]. Genomic analyses of diverse microbes provide similar results [5–10], and it seems that a majority of genes in any pan-genome will be comprised of either high-frequency core genes or low-frequency strain-specific genes [11,12]. These patterns of genomic diversity reveal the fundamental impact of HGT on evolution, and they suggest that bacterial and archaeal genomes comprise a dynamic mosaic of horizontally acquired genes whose frequency fluctuates in the population in response to both **selection** (see Glossary) and **genetic drift**.

HGT presents a range of consequences for microbial genomic diversity. On the one hand, high rates of gene exchange and recombination within species, as mediated by homologous recombination, can act as a cohesive force with the potential to purge nucleotide diversity and prevent diversification [13–15]. While on the other hand, interspecies gene exchange, as mediated by illegitimate recombination, can promote diversification through both the acquisition of niche-expanding adaptive genes [16–18] and the imposition of barriers to homologous recombination [19–21]. HGT also provides a mechanism whereby asexual organisms can realize some of the evolutionary benefits of sexual reproduction [22]. In this way, HGT may allow some microbial lineages to operate in the manner of 'biological species' [23] whose boundaries are described by patterns of gene exchange [13,24]. While microbial genomes are constantly sampling genes from their surroundings, patterns of microbial evolution generally coalesce into tree-like forms [25], and many lineages exhibit sufficient phenetic and phylogenetic coherence to be described as distinct species [26,27]. Hence, the evolutionary consequences of HGT must be governed by constraints which act both at the level of gene frequency and gene transfer.

Trends

Microbial genomic diversity is often explained by invoking selection acting on large populations at demographic equilibrium. However, historical fluctuations in population size can produce nonadaptive changes in genomic diversity due to drift.

Gene surfing can explain patterns of genomic diversity in diverse species of plants and animals as a consequence of postglacial range expansion driven by historical climate change during the Pleistocene.

We propose that genome surfing can result when the demographic mechanisms which produce gene surfing act on microbial populations capable of horizontal gene transfer (HGT).

Patterns of genetic diversity and gene flow within *Streptomyces* are indicative of postglacial demographic range expansion. Genome surfing provides a mechanism to explain ancestral patterns of horizontal gene exchange and current patterns of genomic diversity observed within *Streptomyces*.

[1]School of Integrative Plant Science, Cornell University, Ithaca, NY 14850 USA
[2]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham NH 03824, USA

*Correspondence: dhb28@cornell.edu (D.H. Buckley).

Constraints on gene transfer relate to the likelihood that DNA from one population will co-occur with a cell from a second population, be taken up by that cell, and be stably integrated into its chromosome. The likelihood of gene transfer between cells is ultimately affected by **propinquity** (i.e., the probability of two cells co-occurring in space and time) and genetic compatibility. *Propinquity* is governed by geographic range, ecological niche, dispersal characteristics, and abundance of both donor and recipient populations. Genetic compatibility between donor and recipient is governed by mechanisms of DNA transfer and integration, sequence similarity, codon usage, metabolic compatibility, and restriction systems, among other factors. Constraints on gene frequency, in contrast, relate to the likelihood that a segment of transferred DNA will contribute to the genomic diversity of descendent populations. Hence, gene frequency within a population is controlled by selection and genetic drift.

Gene frequency distributions provide insight into the evolutionary mechanisms that govern genomic diversity by expressing gene content conservation across a given sample of genomes [28]. Microbial gene frequency distributions typically conform to a characteristic 'U-shape' (Figure 1, upper right panel) in which the majority of genes are observed either at very high frequency (the core genes) or very low frequency (unique genes) in the population [7,29]. It is hypothesized that the shape of this distribution results from the balance of gene acquisition and gene loss [12,29–31], yet it is broadly debated whether gene frequency distributions are the result of neutral processes or selection [12,32,33]. Selection produces a strong deletion bias in microbial genomes, causing the loss of genes that lack adaptive benefits [30,34]. The vast majority of genes acquired by HGT are neutral, nearly neutral, or deleterious [35], and these nonadaptive genes are likely transient within genomes and are likely to fluctuate randomly at low frequency within populations. In contrast, adaptive genes once acquired by HGT will increase in frequency within a population. For example, horizontally acquired genes that mediate alginate degradation have caused **selective sweeps**, which have facilitated adaptive radiation within the Vibrionaceae [16,17,35]. In contrast, genes will move from high to low frequency if they no longer convey adaptive benefits, which can occur in response to a change in environmental conditions or a change in the organism's ecological niche.

It is clear that selection has profound impacts on the frequency of horizontally acquired genes and the shape of gene frequency distributions. However, selection cannot readily explain all patterns of HGT. In particular, housekeeping genes often reveal signatures of interspecies or even interphylum gene exchange [36–38], though orthologous replacement of divergent housekeeping genes is unlikely to provide adaptive benefits. Housekeeping genes, which are components of the core genome and encode essential cellular functions, typically have low rates of sequence evolution because maladaptive mutations are purged rapidly by selection and because beneficial mutations are rare. A housekeeping gene that is replaced by interspecies HGT is much more likely to be deleterious than beneficial [39–41]. Indeed, HGT events can impose major costs on the recipient cell, including disruption of genomic features, cytotoxic effects, metabolic costs, and disruption of cellular networks [41,42]. As a consequence, selection would not be expected to favor interspecies HGT of housekeeping genes. Hence, the acquisition of adaptive genes by HGT, while clearly important as a driver of niche expansion and adaptive radiation [16,18], is insufficient to explain the effects of HGT on the genomic diversity of microbial populations. We propose that gene acquisition and loss dynamics vary over time for pan-genomes as a consequence of punctuated equilibria caused by historical changes in demography. Furthermore, we propose ***genome surfing*** as a mechanism to explain how transient changes in demography can have lasting impacts on the pan-genome composition of certain microbes.

Glossary

Bottleneck: a severe reduction in a population size.
Effective population size (N_e): the number of individuals of an idealized population needed to capture the genetic diversity of the actual population.
Expansion edge: the moving boundary of a range expansion.
Founder effects: a reduction in genetic diversity caused when a small number of individuals founds a population.
Gene surfing: a genetic consequence of range expansion driven by neutral variations in gene frequencies that occur at an expansion edge.
Genetic drift: an evolutionary mechanism whereby allele or gene frequencies change as a result of random sampling effects.
Genome surfing: contemporaneous introgression of many horizontally acquired genes into a pan-genome as a consequence of gene surfing.
Introgression: the permanent incorporation of genes from one pan-genome into a second pan-genome of a divergent lineage.
Isolation by distance: a mechanism that causes spatial gradients of genetic diversity when dispersal limitation allows the local accumulation of genetic variation.
Microbial sectoring: a radial pattern of genetic discontinuity produced during range expansions.
Panmixia: a population characterized by unlimited gene flow.
Propinquity: the physical co-occurrence of items in space and time.
Range expansion: an event that occurs when a population colonizes a geographic region which it did not previously occupy, and which may or may not already be occupied by other populations, as facilitated by dispersal.
Selection: an evolutionary mechanism whereby allele or gene frequencies change in response to fitness effects.
Selective sweep: a loss of genetic variation resulting from a rapid fixation of a beneficial allele or gene.

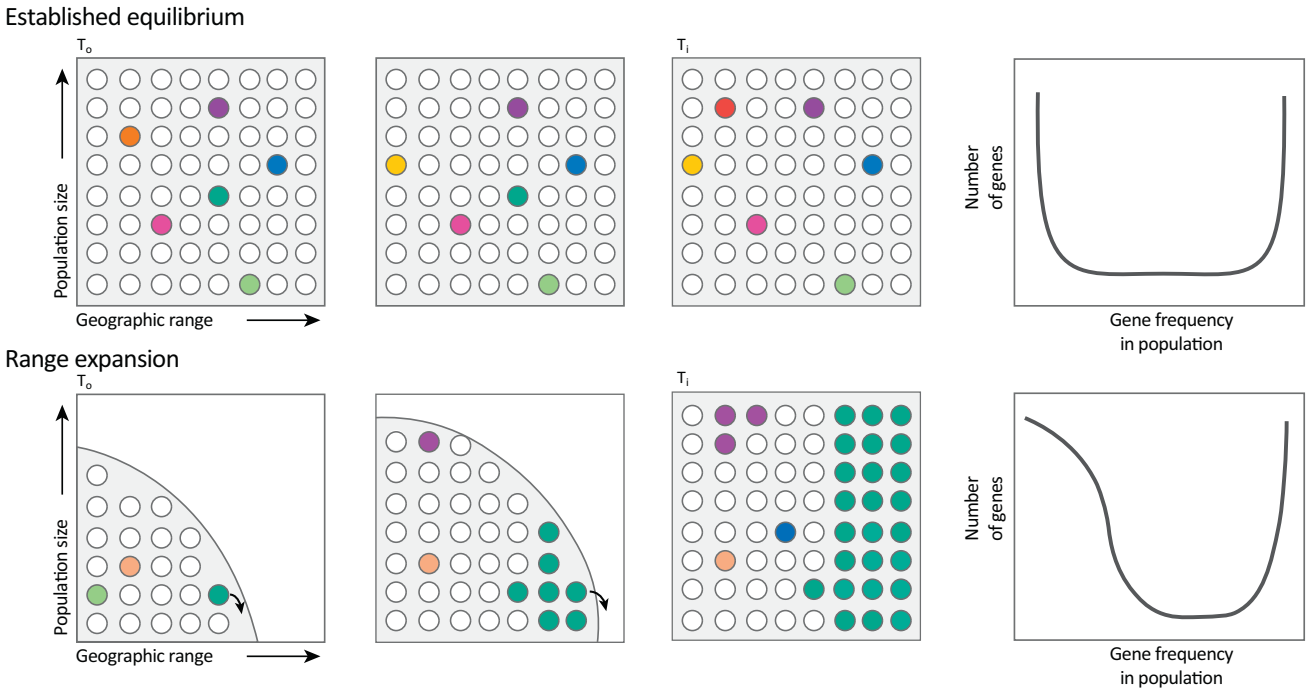


Figure 1. Effect of Demographic Range Expansion and Gene Surfing on Genomic Diversity. At demographic equilibrium (top row) the rate of new genes acquired by horizontal gene transfer (HGT) will be balanced by rates of gene loss due to deletion and drift resulting in a 'U-shaped' gene frequency distribution for the population. Strains that have recently acquired a new gene as a result of HGT are indicated by a change in color, while deletion of that gene is indicated by loss of color. Only recent acquisition events are shown for simplicity. During a demographic range expansion (bottom row), founder effects cause newly colonized sites along the expansion edge to have a small effective population size (N_e) (x-axis). Genes acquired by individuals colonizing new sites (green circles) increase in frequency (y-axis) within these sites as a result of drift. As the expansion edge moves forward additional sites are preferentially colonized by lineages present at the expansion edge (arrows), further amplifying founder effects, and allowing newly acquired genes to increase in frequency across the newly occupied territory. These founder effects will cause an increase in the number of intermediate frequency genes across the full range of the population (bottom right).

Genome Surfing and Microbial Diversification

Widespread evidence for interspecies, and even interphylum, HGT of housekeeping genes suggests that selection pressure may vary over time, with periods of relaxed selection favoring gene acquisition over gene loss. Examples of relaxed selection are common in nature, being caused by *population* **bottlenecks**, demographic expansion, changes in predation pressure, or changes in environmental conditions [43]. The unifying feature of such events is the increased power of genetic drift, which occurs when **effective population size** (N_e) is reduced. Genetic drift is a fundamental evolutionary mechanism that can promote the fixation of neutral and deleterious mutations. For example, the effects of genetic drift are strong in obligate endosymbionts whose N_e is small due to vertical transmission within host populations [44,45]. However, many free-living microbes have a very large N_e, rendering the effect of drift negligible [3,46]. Yet, it is difficult to measure the N_e for microbial populations, and episodic demographic phenomena that cause a temporary reduction in N_e, such as a population bottleneck or demographic **range expansion**, can promote the fixation of nonadaptive genes [47].

The power of demographic range expansion to cause spatially structured patterns of genetic diversity is governed by rates of dispersal and gene flow [47]. High dispersal promotes **panmixia**, causing gene frequency distributions governed by selection acting globally. However, at intermediate levels of dispersal, range expansion can occur gradually in response to changing environmental conditions, such as those caused by changes in climate, habitat, or

land use. Indeed, climate-driven range expansions are well documented for a diversity of plant, animal, and even microbial species [48–55].

Gene surfing can result from demographic range expansion at intermediate levels of dispersal [56]. During a demographic range expansion, a portion of the population will form an **expansion edge** as it spreads into a new range (Figure 1). Individuals along this expansion edge are constantly dispersing to new sites and founding new local subpopulations [56,57]. As a result, **founder effects** have a strong influence on genomic diversity along the expansion edge when the outgrowth of founding colonists exceeds the arrival of new immigrants [47]. These founder effects cause a dramatic reduction in N_e within these newly occupied sites, manifesting as a series of localized population bottlenecks [56,57]. In this way, the expansion edge represents a rolling bottleneck, which spreads across the landscape leaving a genetic legacy of drift in its wake [57]. Gene surfing thereby provides a demographic mechanism whereby the frequency of neutral, and even deleterious genes, can increase dramatically within a population [56].

Movement of the expansion edge across the landscape acts to magnify the impact of genetic drift across large spatial scales (Figure 2). As the edge rolls out across the landscape, the population will exhibit a reduction in allelic diversity parallel to the axis of expansion because of founder effects and because microbes at the expansion edge are more likely to colonize new sites than microbes from the center of the ancestral range [57]. **Microbial sectoring** will further emphasize the ability of drift to structure the population, resulting in geographic diversity gradients that run perpendicular to the axis of expansion [58]. Microbial sectoring occurs when unique genetic changes, which occur at discrete sites along the expansion edge, propagate out across the landscape because of range expansion. Microbial sectoring results in a radial pattern of genetic diversity characterized by genetic discontinuity between adjacent geographic sectors [58]. Over longer evolutionary timescales, dispersal and intraspecies gene exchange decay these local diversity gradients, but the population will continue to manifest many genetic consequences of demographic range expansion and drift [57].

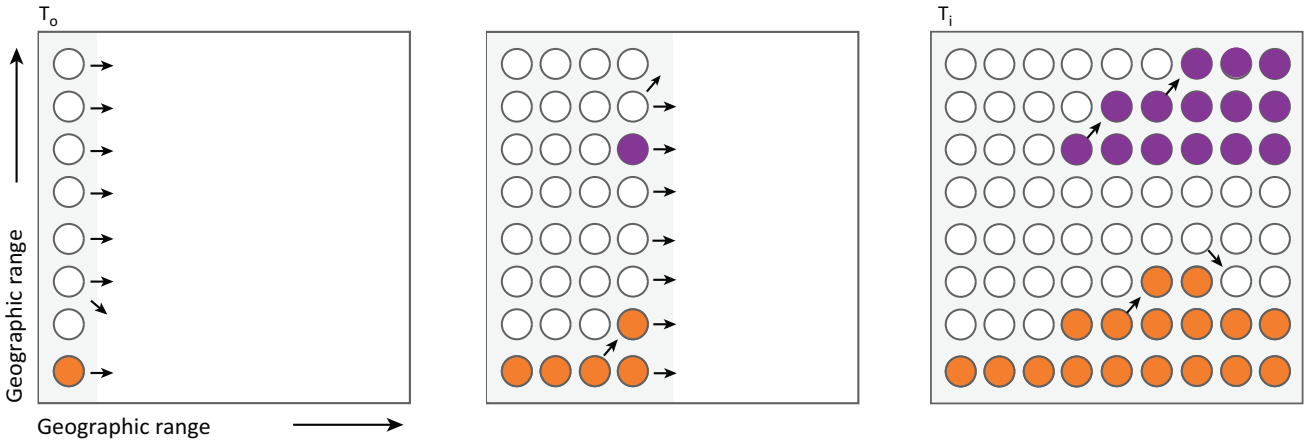
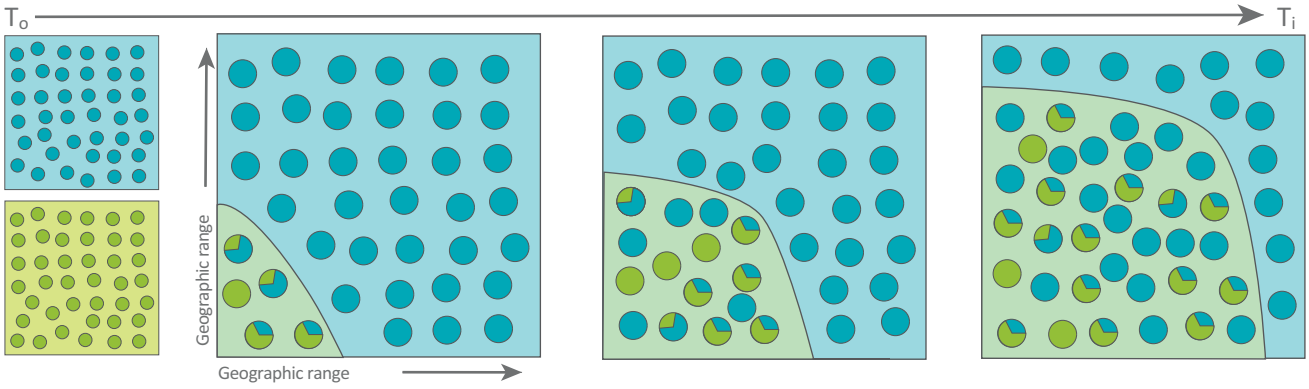


Figure 2. Microbial Sectoring Can Generate Spatial Patterns of Genetic Diversity. During demographic range expansion, individuals at the expansion edge preferentially colonize new sites. Strains that have recently acquired a new gene as a result of horizontal gene transfer (HGT) are indicated by a change in color. Genetic processes (mutations or HGT events) that occur along the expansion edge increase in frequency across space as the edge expands outward. Dispersal along the edge (arrows) is stochastic, resulting in sectoring that causes discontinuous gradients of genetic diversity perpendicular to the axis of expansion [58]. Allelic diversity will decline parallel to the axis of expansion as a result of founder effects that occur along the expansion edge [57].

The mechanism of gene surfing can explain certain patterns of genomic diversity driven by HGT. For example, geographic range expansion facilitates interspecies hybridization and **introgression** when invading species expand into ranges occupied by other species [57,59]. Again, founder effects at the expansion edge amplify this effect. The effects of gene surfing will cause asymmetrical gene exchange between native and invading species (Figure 3) because the native population has a larger and more dense population than the invader [59]. This demographic difference means that both native and invading cells are much more likely to encounter DNA from native species than from invading species. Furthermore, native species are more likely to have a large N_e, causing acquired genes to remain at low frequency, while invading species have a low N_e, increasing the likelihood that drift will cause acquired genes to increase in frequency. Indeed, coalescent modelling of microbial populations predicts that differences in N_e can promote asymmetrical introgression between microbial species [14].

In this way gene surfing can facilitate *genome surfing,* which we define as the contemporaneous introgression of many horizontally acquired genes into a pan-genome driven by demographic phenomena and independent of selection. Simply stated, genome surfing occurs when the genomic consequences of HGT are amplified by the demographic phenomena that produce gene surfing. Genome surfing would differ from gene surfing in at least two major ways. First, genome surfing acts primarily at the level of the microbial pan-genome (i.e., gene presence and absence across genomes) while gene surfing acts at the level of the genome (i.e., allele frequency across genomes). Second, genome surfing can have a much larger impact on genome dynamics than gene surfing since genome surfing acts on a gene pool which, because of HGT, is potentially unlimited in size and diversity, while gene surfing acts on a gene pool defined by the sexual boundaries of a species.

Genome surfing provides a mechanism to explain the presence of horizontally acquired genes which were presumably nonadaptive at the time of transfer. Gene surfing favors the introgression of acquired genes into an expanding population because of relaxed selection and drift [57]. The carriage of nonadaptive genes at high frequency during prolonged periods of relaxed selection can facilitate the development of compensatory mutations, which ameliorate selection costs imposed by horizontally acquired genes [60–64]. Hence, genome surfing provides an evolutionary mechanism through which large numbers of horizontally acquired genes can



Trends in Microbiology

Figure 3. Demographic Range Expansion Can Promote Genome Surfing of Horizontally Acquired Genes. Demographic range expansion causes founder events along the expansion edge (see Figure 1). Asymmetrical introgression of horizontally acquired genes will be favored by differences in effective population size (N_e) between native (blue, high N_e) and expanding (green, low N_e) populations. Low N_e along the expansion edge promotes relaxed selection and gene surfing, thereby allowing introgression of horizontally acquired genes into the expanding population [57]. While bidirectional gene exchange can occur, the high N_e of the native population will block introgression. As a result, demographic range expansion can cause a large number of horizontally acquired genes to rapidly increase in frequency in the expanding population.

increase in frequency within the pan-genome, and even enter the core genome of a microbial species, independent of selection. Furthermore, microbial sectoring allows for the geographic expansion of these introgressed genomes across the landscape resulting in patterns of genomic diversity correlated with, but not caused by, environmental gradients. Eventually, dispersal and intrapopulation gene exchange will bring the population to equilibrium across its new range, thereby enabling selection to purge nonadaptive genes acquired during expansion. However, compensatory mutations will facilitate the retention of genes acquired by genome surfing. In addition, if dispersal is limiting across the new spatial range then genomic diversity introduced by genome surfing will reinforce lineage subdivision and divergence.

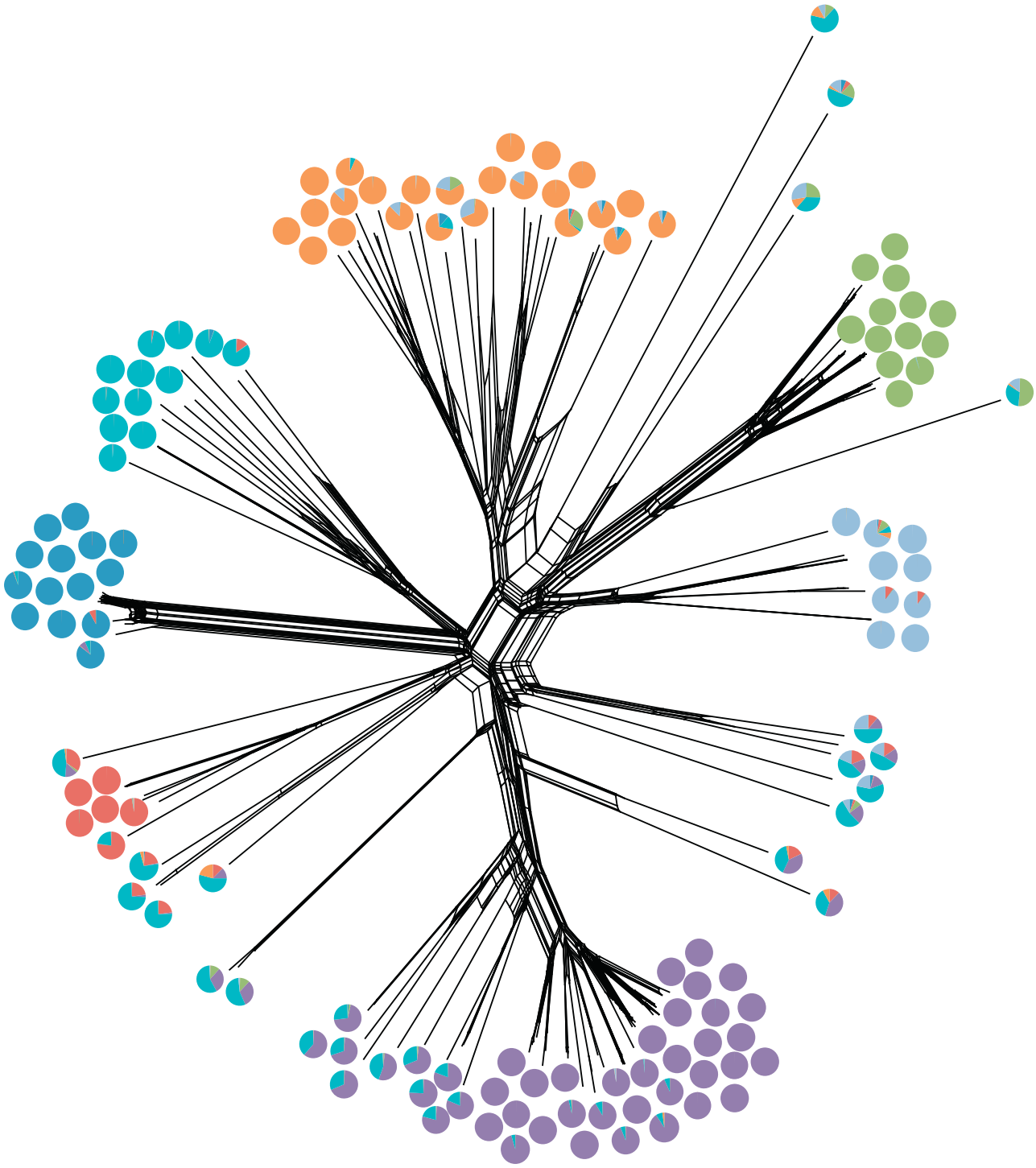
Genome surfing also provides a hypothesis to explain how microbial genomes can increase in size. While several mechanisms promote genome streamlining [3,45,65], fewer hypotheses explain genome expansion [66]. Currently, the most common hypothesis to explain the existence of large microbial genomes is that large genomes convey metabolic versatility [67,68]. Genome surfing provides a mechanism whereby many newly acquired genes can fix contemporaneously, facilitating an increase in genome size as a consequence of both gene acquisition and relaxed selection. Hence, genome surfing predicts genome expansion as a consequence of historical demographic phenomena. In this way, large genome size in bacteria may be a spandrel, a feature that arises not as a result of adaptation, but rather as a byproduct of some other evolutionary process [69]. In this case, a large genome size may arise as a byproduct of historical demographic phenomena.

In Practice, Exploring *Streptomyces* Biogeography

Under what conditions might we find evidence for genome surfing in microorganisms? First, genome surfing would be most likely to occur in populations that experience intermediate levels of dispersal. Human pathogens and marine organisms typically exhibit high rates of dispersal, making genome surfing unlikely. Contiguous terrestrial habitats, however, allow for dispersal limited by distance which would be expected to facilitate genome surfing. Second, the genetic consequences of range expansions are most likely to occur in response to large-scale changes in habitat, such as those caused by climate change. Third, on evolutionary timescales, the effects of genome surfing will be most evident during or soon after the expansion event, but before dispersal has a chance to weaken diversity gradients formed during expansion. Fourth, we would predict evidence for genome surfing in microbes whose genomes have increased in size dramatically relevant to a recent common ancestor. Finally, because the genetic consequences of range expansions result from neutral processes, detection of genome surfing requires that we examine genetic diversity using approaches that are sensitive to genetic drift. The best approach to examine the effects of genetic drift on patterns of microbial diversity is to focus on closely related populations or species which are easily cultivated, occur in similar ecological habitats, and are spread across a wide geographic range [70].

The genus *Streptomyces* provides an excellent system for investigating how dispersal, genetic drift, and gene exchange generate patterns of genomic diversity. *Streptomyces* species are ubiquitous bacteria in soil, and they are cultivated readily on laboratory media. *Streptomyces* species are important contributors to ecosystem function in terms of the degradation of complex plant polymers, such as cellulose and lignin [71], but they are best known for their ability to produce a vast repertoire of secondary metabolites, including a majority of antibiotics in use clinically [72]. Furthermore, the evolution of these secondary metabolite gene clusters has been influenced by HGT [73,74].

Streptomyces genomes display evidence of widespread HGT both within and between species indicative of reticulate evolution (Figure 4). Within *Streptomyces* species, rates of intraspecies gene exchange and recombination are among the highest observed for any species of bacteria



Trends in Microbiology

Figure 4. *Streptomyces* Exhibit a Reticulate Phylogeny. Neighbor net phylogeny/tree constructed using six housekeeping genes from strains representing diverse species of *Streptomyces*, reproduced from [38]. Interspecies horizontal gene transfer (HGT) events generate conflicting phylogenetic signals, and these conflicts are interpreted by addition of edges to the network diagram. Pie charts associated with each strain indicate patterns of genetic ancestry inferred from STRUCTURE analysis of these six genes [95]. STRUCTURE identified seven ancestral populations whose contribution to the ancestry of each strain is indicated with different colors representing each ancestral population. *Streptomyces* species form distinct species clusters, and mosaic patterns of ancestry can easily be detected across the six

(Figure legend continued on the bottom of the next page.)

[37,38,75]. Furthermore, housekeeping genes are often incongruent between species, and statistical tests for recombination suggest that many species have acquired core genes as a result of interspecies HGT. *Streptomyces* species have physiological traits that may facilitate high rates of HGT. For example, *Streptomyces* species have an unusual conjugation mechanism which mediates double-stranded DNA transfer [76], mobilizes chromosomal markers at high frequency [77,78], and which also allows interspecies hybridization and the formation of hybrid strains containing equivalent genetic contributions from divergent species [79,80]. It is clear that *Streptomyces* species possess mechanisms of gene transfer that permit widespread HGT within and between species; however, the evolutionary mechanisms that cause horizontally acquired genes to occur at high frequency within *Streptomyces* populations are less discernable.

We investigated patterns of genetic diversity across spatial scales within *Streptomyces* by generating a culture collection of strains isolated from soils across the United States. These strains were all isolated from similar habitats, comprising slightly acidic arable soils associated with perennial grasses [81], thereby minimizing adaptive differences imposed by selection acting across different habitats and enhancing our ability to observe the effects of drift [70]. Genetic dissimilarity of communities across sites was correlated with geographic distance but poorly correlated with environmental dissimilarity, and this result is consistent with dispersal limitation [81]. Multilocus sequence analysis (MLSA) of species-like populations reveals that while identical alleles were found at low frequency in sites separated by more than 6000 km, all but one of the concatenated MLSA haplotypes were site-specific [82]. This observation is consistent with intermediate levels of dispersal across the observed range. That is, limits to dispersal restrict MLSA haplotypes to local populations (i.e., private haplotypes), but dispersal is sufficient to prevent the fixation of private alleles (i.e., alleles unique to a single subpopulation) in local populations. *Streptomyces* species have a unique developmental cycle characterized by growth of filamentous mycelia followed by the formation of aerial hyphae containing arthrospores. The function of arthrospores appears to be desiccation resistance and dispersal, and spores are dispersed readily by wind and animals [83,84]. Hence, while *Streptomyces* species have the capacity for long-range dispersal, patterns of genetic diversity reveal dispersal limitation acting at regional spatial scales.

Patterns of genetic diversity and population structure within our *Streptomyces* strain collection are consistent with the hypothesis that these patterns are a result of historical demographic phenomena caused by oscillations in climate [81,82]. In particular, species diversity declines with increasing latitude, and northern lineages are phylogenetically clustered and have lower mean phylogenetic root depth than southern lineages [81]. Furthermore, patterns of gene flow partition with latitude, consistent with the maximum extent of glaciation, and intraspecies genetic diversity also correlates with latitude [82]. These observations are consistent with the conclusion that glaciation during the late Pleistocene prevented colonization of northern sites, and that glacial retreat promoted subsequent northward range expansion. Similar paleological historic demographic events have been associated with gene surfing in a range of organisms [85–87].

We observe additional evidence of recent demographic expansion at the species level in *Streptomyces*. The association of alleles at different loci, or linkage disequilibrium (LD), can be used to assess recombination and population structure in microbes [88]. LD can be evaluated in a statistical framework using the Index of Association (I_A), where $I_A = 0$ indicates a random association of alleles, known as linkage equilibrium, and values greater than 1 indicate

housekeeping genes analyzed. Mosaic patterns tend to be shared by members of species clusters, suggesting that these patterns are caused by evolutionary events that occurred in the most recent common ancestor of these clades.

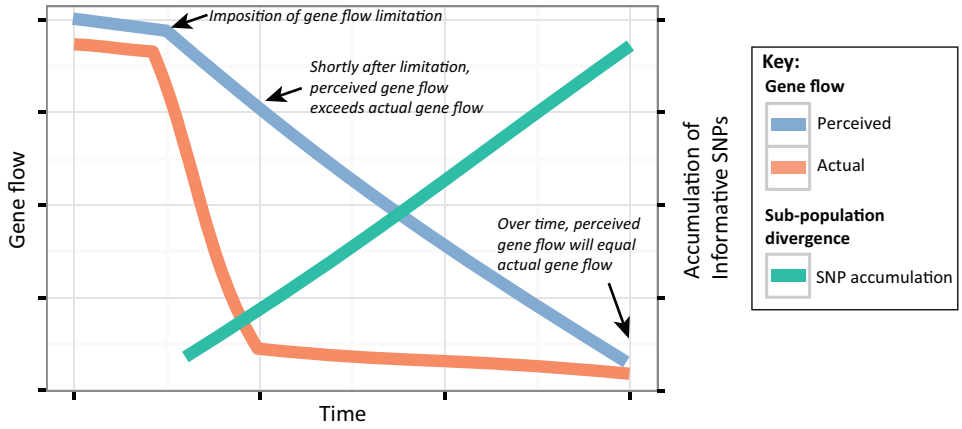


Figure 5. Detection of Dispersal Limitations Requires Polymorphisms Which Accumulate over Time. Range expansion can produce **isolation by distance**, but if the expansion event occurred in recent evolutionarily history then the accumulation of mutations may be insufficient to detect contemporary rates of dispersal and gene flow between spatial demes. Spatially explicit single-nucleotide polymorphisms (SNPs), defined here as SNPs that reveal spatial patterns, increase gradually over time due to mutation. Isolation by distance or vicariance will reduce actual gene flow rates, but perceived gene flow greatly exceeds actual gene flow when the number of spatially explicit SNPs is low, possibly leading to an incorrect conclusion of panmixia. Hence, a lack of spatially explicit SNPs in a population that spans a given range could result either from very high rates of recombination across that range or be due to an evolutionarily recent geographic range expansion.

population structure. We observe evidence for linkage equilibrium in *Streptomyces* species whose current ranges span the historical boundaries of glaciation. For example, *S. pratensis*, found in North Carolina, New York, Michigan, and Quebec, has $I_A = 0.0018$ [37]. Likewise, *Streptomyces* phylogroup MAN125 found in the Pacific Northwest and Alaska has $I_A = 0.09$ [82]. Linkage equilibrium across the geographic range of a species is consistent with contrasting scenarios: (i) a panmictic population lacking contemporary barriers to gene flow, or (ii) a recent demographic range expansion [89]. The latter case results when contemporary dispersal and gene flow are actually limited by geographic distance, but the genetic effects of restricted gene flow are not yet perceived because the accumulation of spatially explicit polymorphisms is insufficient to differentiate between extant subpopulations (Figure 5). In this way, a recent range expansion can generate patterns of genetic diversity that suggest panmixia despite the contemporary presence of geographic barriers to gene exchange. Given the evidence for *Streptomyces* dispersal limitation [81,82] and the observation of linkage equilibrium in *S. pratensis* [37] and MAN125 [82], it is parsimonious to conclude that these populations have each experienced a northward range expansion in recent evolutionary time.

Genome Surfing and Ancestral Patterns of HGT

Many ecological phenomena facilitate range expansions, but those caused by historical climate change have left their legacy on the genetic diversity of diverse plants, animals, and microbes [47–49,90,91]. In particular, glacial periods have profound effects on soil organisms. Glaciers scrape off the surface of the Earth as they advance, removing soil and leaving gouges in exposed stone. Glaciers also eliminate plants, which are a dominant source of organic matter sustaining most soil organisms, including *Streptomyces*. Glacial periods have reoccurred roughly every 100 000 years over the last 800 000 years [92]. During these glacial periods, glaciers eliminated organisms from the soil surface for thousands or tens of thousands of years. During interglacial periods, glacial retreat allows northward range expansion and subsequent regional diversification of populations inhabiting (mostly southern) glacial refugia. The arrival of the next glacial period then drives these northern populations to (mostly) southern ice-free

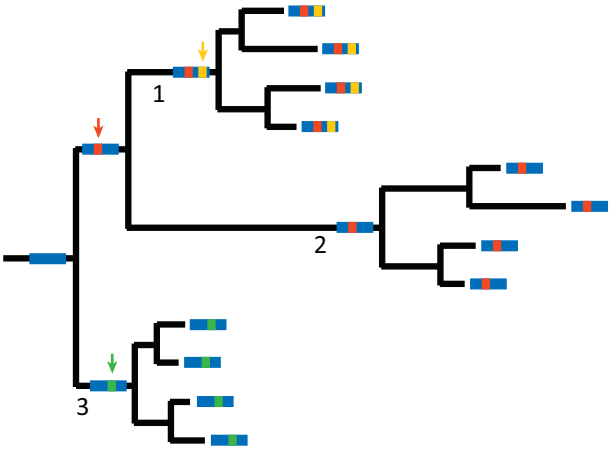


Figure 6. Discrete Horizontal Gene Transfer (HGT) Events Which Occur in Ancestral Nodes Can Cause Widespread and Complex Patterns of Mosaicism in Extant Lineages. A small number of HGT events (indicated by colored arrows), which occur in ancestral nodes, cause reticulate ancestry in all extant lineages (numbered 1–3). Note that undersampling of the phylogeny will obscure the ancestral origin of these three gene transfer events, likely leading to the erroneous conclusion that gene exchange was driven by contemporary processes such as panmixia.

refugia once more where they must compete against their ancestors or go extinct. Hence, glacial cycles have provided numerous opportunities for repeated periods of both northward and southward range expansion over millions of years. If range expansion promotes genome surfing, then current patterns of genomic diversity may represent the historical legacy of repeated waves of demographic expansion as driven by historical patterns of climate change. Repeated periods of genome surfing could generate reticulate phylogenies, such as those observed for *Streptomyces* (Figure 5), in which the core genome displays mosaic ancestry comprised of genes derived from multiple ancestral species.

We propose that the reticulate evolutionary history of *Streptomyces* is a product of genome surfing driven by repeated periods of range expansion precipitated by changes in paleoclimate. We have evidence that reticulate evolution in *Streptomyces* is a consequence of ancestral rather than contemporary processes. For example, analysis of diverse *Streptomyces* species revealed that 71% of strains had mosaic ancestry with genes derived from different ancestral species [38]. However, most (88%) of these genes are monophyletic at the taxonomic level of species clusters and their subclades, suggesting that these HGT events are derived from a far smaller number of interspecies HGT events which occurred in ancestral lineages. That is, most inferred interspecies HGT events within the *Streptomyces* core genome were the result of historical HGT events, rather than contemporary patterns of gene exchange. The fixation of HGT events in ancestral lineages can produce contemporary lineages that have inherited 'horizontally acquired' genes through vertical descent (Figure 6). While the ancestral acquisition and fixation of HGT events can be driven by both neutral and adaptive processes, the majority of HGT events are neutral or near-neutral [35]. In addition, there is little evidence that interspecies HGT of housekeeping genes provides adaptive benefits, and there are several lines of evidence which suggest that such transfers are neutral or deleterious. Genome surfing provides a plausible, though not exclusive, mechanism that can explain the contemporaneous ancestral fixation of numerous nonadaptive or maladaptive genes acquired by HGT.

The ancestral fixation of genes acquired by HGT, and the use of these fixation events to understand phylogenesis, is not a new concept. For example, the presence of Chlamydiae-related genes in red algae, glaucophytes, and green plants suggest that HGT occurred once in the common ancestor of these primary photosynthetic eukaryotes [93]. Hence, when a horizontally transferred gene persists in a lineage, it represents an important landmark which can be used as a derived character to reconstruct phylogenetic relationships and evolutionary processes [94].

Rates and patterns of gene flow can vary tremendously during the evolution of a microbial lineage in response to habitat change over time. Most attempts to quantify gene flow, however, focus on contemporary rather than historical processes and often make the explicit assumption that populations have maintained demographic equilibrium over their evolutionary history. An understanding of historical processes can therefore have dramatic effects on the manner in which we interpret reticulate evolutionary patterns. For example, repeated range expansions caused by glacial cycles over the last 800 000 years, and during prior ice ages, could have driven serial genome surfing events, resulting in the widespread and complex patterns of genetic mosaicism we observe in *Streptomyces* (Figure 5). Note that historical interpretations of gene exchange in no way preclude the presence or importance of contemporary gene-exchange processes. They are simply an evolutionary manifestation of nonequilibrium gene exchange processes playing out over time.

Concluding Remarks

Genome surfing provides a compelling neutral mechanism to explain patterns of genomic diversity. While selection will act to refine genome composition in response to contemporary conditions, genome surfing offers a historical mechanism that can influence the underlying genomic variation upon which selection acts. Predictions of genome surfing are readily testable through comparative genomic analysis of spatially structured microbial populations (see Outstanding Questions). For example, we expect to observe: (i) allelic diversity to decline along the axis of an expansion event, (ii) microbial sectoring to produce gradients of diversity perpendicular to the axis of expansion, (iii) pan-genome frequency distributions to reflect elevated levels of intermediate frequency genes acquired by HGT, (iv) genetic evidence of relaxed selection and a reduction in N_e (v), and an increase in genome size in derived lineages relative to their recent ancestors. As microbiologists, we have a natural desire to understand the genetic basis of ecological phenomena which often causes us to favor adaptive explanations for contemporary patterns of genomic diversity, but it is useful to remember that historical contingency is common throughout evolutionary history. Just ask the dinosaurs; whether meteor or glacier, discrete historical events can have major evolutionary consequences. Ultimately, we expect that a range of evolutionary mechanisms govern extant patterns of genomic diversity, including both contemporary adaptive processes as well as historical demographic processes. Furthermore, these evolutionary mechanisms may vary in their relevance for different types of microbes such as those from marine, terrestrial, and pathogenic systems. We argue that a greater understanding of the impacts of historical demography will improve our ability to explain microbial genomic diversity and mechanisms of phylogenesis.

Outstanding Questions

Do microorganisms in terrestrial environments truly have distinct spatial ranges governed by dispersal limitation, or is dispersal unlimited?

Are there limits to microbial gene flow in terrestrial environments, and if so, what are the limits and at what spatial scales do they operate?

To what degree do patterns of dispersal limitation and gene flow vary between microbial species?

Can current patterns of genomic diversity be explained as a function of historical range expansions as predicted by gene surfing and microbial sectoring?

To what degree does genome surfing in *Streptomyces* contribute to the diversification of secondary metabolite gene clusters?

Can increases in genome size be produced as a consequence of relaxed selection in response to bottleneck effects caused by demographic range expansion?

To what degree is microbial evolution governed by punctuated equilibrium, resulting from historical changes in climate?

References

1. Perna, N.T. *et al.* (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529–533
2. Welch, R.A. *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 17020–17024
3. Giovannoni, S.J. *et al.* (2014) Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1553–1565
4. Kaas, R.S. *et al.* (2012) Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* 13, 577
5. Tettelin, H. *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13950–13955
6. Grote, J. *et al.* (2012) Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* 3, e00252-12
7. Lapierrel, P. and Gogarten, J.P. (2009) Estimating the size of the bacterial pan-genome. *Trends Genet.* 25, 107–110
8. Loper, J.E. *et al.* (2012) Comparative genomics of plant-associated pseudomonas spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet.* 8, e1002784
9. Zhaxybayeva, O. *et al.* (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5865–5870
10. Fouts, D.E. *et al.* (2008) Complete genome sequence of the N-2-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS Genet.* 4, e1000141
11. Medini, D. *et al.* (2005) The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594
12. Lobkovsky, A.E. *et al.* (2013) Gene frequency distributions reject a neutral model of genome evolution. *Genome Biol. Evol.* 5, 233–242
13. Fraser, C. *et al.* (2007) Recombination and the nature of bacterial speciation. *Science* 315, 476–480
14. Doroghazi, J.R. and Buckley, D.H. (2011) A neutral model for the effect of homologous recombination on microbial diversification. *Genome Biol. Evol.* 3, 1349–1356

15. Didelot, X. *et al.* (2012) Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* 13, 256
16. Hehemann, J.H. *et al.* (2016) Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nat. Commun.* 7, 12860
17. Shapiro, B.J. *et al.* (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* 336, 48–51
18. Cordero, O.X. and Polz, M.F. (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12, 263–273
19. Vetsigian, K. *et al.* (2011) Structure and evolution of *Streptomyces* interaction networks in soil and *in silico*. *PLoS Biol.* 9, e1001184–e1001184
20. Zawadzki, P. *et al.* (1995) The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. *Genetics* 140, 917–932
21. Dykhuizen, D.E. and Green, L. (1991) Recombination in *Escherichia-coli* and the definition of biological species. *J. Bacteriol.* 173, 7257–7268
22. Moradigaravand, D. and Engelstadter, J. (2013) The evolution of natural competence: disentangling costs and benefits of sex in bacteria. *Am. Nat.* 182, E112–E126
23. Mayr, E. (1982) Speciation and macroevolution. *Evolution* 36, 1119–1132
24. Hanage, W.P. *et al.* (2006) The impact of homologous recombination on the generation of diversity in bacteria. *J. Theor. Biol.* 239, 210–219
25. Puigbo, P. *et al.* (2013) Seeing the Tree of Life behind the phylogenetic forest. *BMC Biol.* 11, 46
26. Kim, M. *et al.* (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351
27. Stackebrandt, E. *et al.* (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047
28. Medini, D. *et al.* (2008) Microbiology in the post-genomic era. *Nature Rev. Microbiol.* 6, 419–430
29. Gordienko, E.N. *et al.* (2013) Evolution of pan-genomes of *Escherichia coli, Shigella* spp., and *Salmonella enterica*. *J. Bacteriol.* 195, 2786–2792
30. Kunin, V. and Ouzounis, C.A. (2003) The balance of driving forces during genome evolution in prokaryotes. *Genome Res.* 13, 1589–1594
31. Berg, O.G. and Kurland, C.G. (2002) Evolution of microbial genomes: sequence acquisition and loss. *Mol. Biol. Evol.* 19, 2265–2276
32. Baumdicker, F. *et al.* (2012) The infinitely many genes model for the distributed genome of bacteria. *Genome Biol. Evol.* 4, 443–456
33. Haegeman, B. and Weitz, J.S. (2012) A neutral theory of genome evolution and the frequency distribution of genes. *BMC Genomics* 13, 196
34. Ochman, H. and Moran, N.A. (2001) Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* 292, 1096–1098
35. Gogarten, J.P. and Townsend, J.P. (2005) Horizontal gene transfer, genome innovation and evolution. *Nature Rev. Microbiol.* 3, 679–687
36. Nesbo, C.L. *et al.* (2006) Recombination in *Thermotoga*: implications for species concepts and biogeography. *Genetics* 172, 759–769
37. Doroghazi, J.R. and Buckley, D.H. (2010) Widespread homologous recombination within and between *Streptomyces* species. *ISME J.* 4, 1136–1143
38. Andam, C.P. *et al.* (2016) Contributions of ancestral inter-species recombination to the genetic diversity of extant *Streptomyces* lineages. *ISME J.* 10, 1731–1741
39. Lind, P.A. *et al.* (2010) Compensatory gene amplification restores fitness after inter-species gene replacements. *Mol. Microbiol.* 75, 1078–1089
40. Wolf, Y.I. *et al.* (2006) Unifying measures of gene function and evolution. *Proc. Roy. Soc. B-Biol. Sci.* 273, 1507–1515
41. Sorek, R. *et al.* (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318, 1449–1452
42. Baltrus, D.A. (2013) Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.* 28, 489–495
43. Lahti, D.C. *et al.* (2009) Relaxed selection in the wild. *Trends Ecol. Evol.* 24, 487–496
44. Funk, D.J. *et al.* (2001) Intraspecific variation in symbiont genomes: bottlenecks and the aphid-buchnera association. *Genetics* 157, 477–489
45. Moran, N.A. and Baumann, P. (2000) Bacterial endosymbionts in animals. *Curr. Opin. Microbiol.* 3, 270–275
46. Batut, B. *et al.* (2014) Reductive genome evolution at both ends of the bacterial population size spectrum. *Nature Rev. Microbiol.* 12, 841–850
47. Waters, J.M. *et al.* (2013) Founder takes all: density-dependent processes structure biodversity. *Trends Ecol. Evol.* 28, 78–85
48. Kenefic, L.J. *et al.* (2009) Pre-Columbian origins for North American anthrax. *PLoS One* 4, e4813
49. Eisenlord, S.D. *et al.* (2012) Dispersal limitation and the assembly of soil Actinobacteria communities in a long-term chronosequence. *Ecol. Evol.* 2, 538–549
50. Mikheyev, A.S. *et al.* (2008) Phylogeography of post-Pleistocene population expansion in a fungus-gardening ant and its microbial mutualists. *Mol. Ecol.* 17, 4480–4488
51. Soltis, D.E. *et al.* (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Syst. Evol.* 206, 353–373
52. Bernatchez, L. and Wilson, C.C. (1998) Comparative phylogeography of nearctic and palearctic fishes. *Mol. Ecol.* 7, 431–452
53. Conroy, C.J. and Cook, J.A. (2000) Phylogeography of a post-glacial colonizer: *Microtus longicaudus* (Rodentia: Muridae). *Mol. Ecol.* 9, 165–175
54. Mila, B. *et al.* (2007) Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus Junco. *Proc. Roy. Soc. B-Biol. Sci.* 274, 2653–2660
55. Wilson, A.B. and Veraguth, I.E. (2010) The impact of Pleistocene glaciation across the range of a widespread European coastal species. *Mol. Ecol.* 19, 4535–4553
56. Edmonds, C.A. *et al.* (2004) Mutations arising in the wave front of an expanding population. *Proc. Natl. Acad. Sci. U. S. A.* 101, 975–979
57. Excoffier, L. *et al.* (2009) Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* 40, 481–501
58. Hallatschek, O. *et al.* (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proc. Natl. Acad. Sci. U. S. A.* 104, 19926–19930
59. Currat, M. *et al.* (2008) The hidden side of invasions: massive introgression by local genes. *Evolution* 62, 1908–1920
60. Lawrence, J.G. and Ochman, H. (1997) Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44, 383–397
61. Bouma, J.E. and Lenski, R.E. (1988) Evolution of a bacteria plasmid association. *Nature* 335, 351–352
62. Lercher, M.J. and Pal, C. (2008) Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.* 25, 559–567
63. Amoros-Moya, D. *et al.* (2010) Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. *Mol. Biol. Evol.* 27, 2141–2151
64. Asai, T. *et al.* (1999) An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1971–1976
65. Lynch, M. (2006) Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* 60, 327–349
66. Cordero, O.X. and Hogeweg, P. (2009) The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21748–21753

67. Konstantinidis, K.T. and Tiedje, J.M. (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3160–3165
68. Daubin, V. and Voran, N.A. (2004) Comment on "The origins of genome complexity". *Science* 306, 978
69. Gould, S.J. and Lewontin, R.C. (1979) The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. *Proc. Roy. Soc. London Ser. B. Biol. Sci.* 205, 581–598
70. Hanson, C.A. *et al.* (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Rev. Microbiol.* 10, 497–506
71. Williams, S.T. *et al.* (1983) Numerical classification of *Streptomyces* and related genera. *J. Gen Microbiol.* 129, 1743–1813
72. Watve, M.G. *et al.* (2001) How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* 176, 386–390
73. Osbourn, A. (2010) Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* 26, 449–457
74. Fischbach, M.A. *et al.* (2008) The evolution of gene collectives: how natural selection drives chemical innovation. *Proc. Natl. Acad. Sci. U. S. A.* 105, 4601–4608
75. Doroghazi, J.R. and Buckley, D.H. (2014) Intraspecies comparison of *Streptomyces pratensis* genomes reveals high levels of recombination and gene conservation between strains of disparate geographic origin. *BMC Genomics* 15, 970
76. Kataoka, M. *et al.* (1991) Regulation and function of the *Streptomyces* plasmid psn22 genes involved in pock formation and inviability. *J. Bacteriol.* 173, 7975–7981
77. Hopwood, D.A. *et al.* (1985) Conjugative sex plasmids of *Streptomyces*. *Basic Life Sci.* 30, 615–634
78. Chater, K.F. *et al.* (1982) Gene cloning in *Streptomyces*. *Curr. Top. Microbiol. Immunol.* 96, 69–95
79. Lomovskaya, N.D. *et al.* (1977) Construction and properties of hybrids obtained in interspecific crosses between *Streptomyces coelicolor* A3(2) and *Streptomyces griseus* Kr15. *J. Gen. Microbiol.* 98, 187–198
80. Alacevic, M. (1963) Interspecific recombination in *Streptomyces*. *Nature* 197, 1323
81. Andam, C.P. *et al.* (2016) A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *mBio* 7, e02200–e02215
82. Choudoir, M. *et al.* (2016) Latitude delineates patterns of biogeography in terrestrial *Streptomyces*. *Environ. Microbiol.* 18, 4931–4945
83. Gorny, R.L. *et al.* (2003) Release of *Streptomyces albus* propagules from contaminated surfaces. *Environ. Res.* 91, 45–53
84. Ensign, J.C. (1978) Formation, properties, and germination of actinomycete spores. *Annu. Rev. Microbiol.* 32, 185–219
85. Neiva, J. *et al.* (2010) Surfing the wave on a borrowed board: range expansion and spread of introgressed organellar genomes in the seaweed *Fucus ceranoides* L. *Mol. Ecol.* 19, 4812–4822
86. Assis, J. *et al.* (2014) Climate-driven range shifts explain the distribution of extant gene pools and predict future loss of unique lineages in a marine brown alga. *Mol. Ecol.* 23, 2797–2810
87. Qi, X.S. *et al.* (2012) Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). *New Phytol.* 196, 617–630
88. Smith, J.M. *et al.* (1993) How clonal are bacteria? *Proc. Natl. Acad. Sci. U. S. A.* 90, 4384–4388
89. Rogers, A.R. (2014) How population growth affects linkage disequilibrium. *Genetics* 197, 1329–1341
90. Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Phil. Trans. Roy. Soc. Lond. Ser. B-Biol. Sci.* 359, 183–195
91. Hewitt, G.M. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* 58, 247–276
92. Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913
93. Huang, J.L. and Gogarten, J.P. (2007) Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* 8, R99
94. Fournier, G.P. *et al.* (2009) Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. *Phil. Trans. Roy. Soc. B-Biol. Sci.* 364, 2229–2239
95. Falush, D. *et al.* (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587