

## Beyond biogeographic patterns: processes shaping the microbial landscape

China A. Hanson<sup>1</sup>, Jed A. Fuhrman<sup>2</sup>, M. Claire Horner-Devine<sup>3</sup> and Jennifer B. H. Martiny<sup>1</sup>

**Abstract** | Recently, microbiologists have established the existence of biogeographic patterns among a wide range of microorganisms. The focus of the field is now shifting to identifying the mechanisms that shape these patterns. Here, we propose that four processes — selection, drift, dispersal and mutation — create and maintain microbial biogeographic patterns on inseparable ecological and evolutionary scales. We consider how the interplay of these processes affects one biogeographic pattern, the distance–decay relationship, and review evidence from the published literature for the processes driving this pattern in microorganisms. Given the limitations of inferring processes from biogeographic patterns, we suggest that studies should focus on directly testing the underlying processes.

### Richness

The number of taxa in a sample, assemblage or community.

### Composition

The identity and relative abundance of taxa in a sample, assemblage or community.

### Taxonomic resolution

The level of genetic variation of the taxa considered.

<sup>1</sup>*Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697 USA.*

<sup>2</sup>*Department of Biological Sciences, University of Southern California, MC0371, Los Angeles, California 90089, USA.*

<sup>3</sup>*School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington 98195, USA. Correspondence to C.A.H. and J.B.H.M.*

*e-mails: [cahanson@uci.edu](mailto:cahanson@uci.edu); [jmartiny@uci.edu](mailto:jmartiny@uci.edu)*

doi:10.1038/nrmicro2795

Published online 14 May 2012

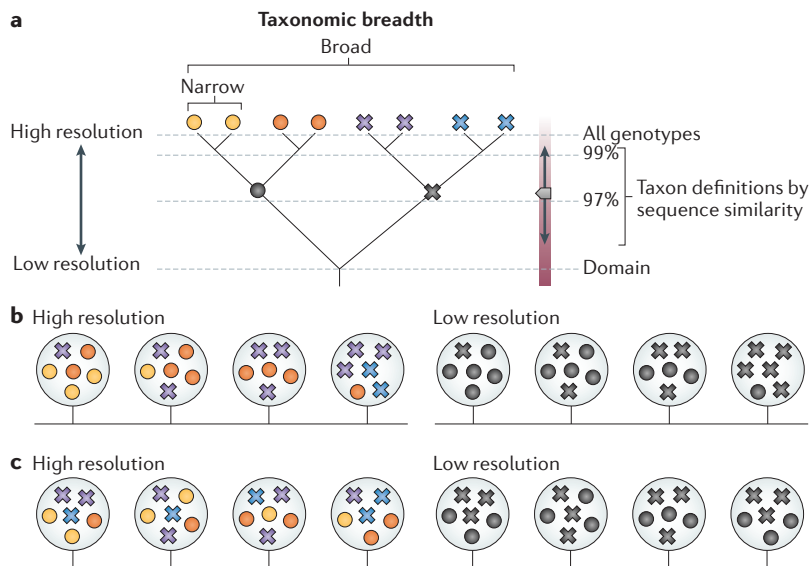
Biogeography is the study of the distribution of organisms across space and time<sup>1</sup>. Traditionally, the field has investigated plants and animals; however, numerous recent studies have focused on the biogeography of microorganisms, including bacteria, archaea, viruses, fungi and other microbial eukaryotes. This growing body of research provides overwhelming evidence that microorganisms display biogeographic patterns, some of which are similar to those of larger organisms (reviewed in REFS 2–6).

Although the existence of microbial biogeographic patterns is now well established, little is understood about the processes underlying them. A central goal of the field of biogeography is to understand the mechanisms that generate and maintain diversity, including its richness and composition. As a result, many microbial studies now focus on what biogeographic patterns reveal about the processes that drive them.

There are several reasons why our understanding of the processes driving microbial biogeographic patterns remains confounded. First, there are numerous theoretical frameworks that researchers use to interpret microbial biogeographic patterns, and the different terminologies and analytical approaches complicate cross-study comparisons. Furthermore, each framework emphasizes a slightly different subset of potential underlying processes, not all of which are considered in each study.

Second, microbial biogeography studies frequently use molecular genetic methods, allowing diversity to be examined along a continuous scale of taxonomic resolution (FIG. 1a). Microbial taxa are then classified into operational taxonomic units (OTUs), which are defined by the nucleotide sequence similarity of one or more genomic regions. Thus, a taxon can be defined by anything from unique nucleotide sequences to groups of sequences that lump together millions of years of evolved diversity (FIG. 1a). Not only does this flexibility make comparisons across studies difficult, it also complicates the interpretation of the processes driving biogeographic patterns.

Finally, as with any field of scientific investigation, there is only so much one can derive about mechanistic processes solely from biogeographic patterns. Indeed, multiple processes can interact in different ways to result in the same pattern. Thus, it is crucial to recognize the limits of what biogeographic patterns can tell us and to supplement work in this area with other approaches that directly assess the underlying processes. To begin to address these challenges, here we consider concepts from population genetics and community ecology to expand the current framework of microbial biogeography, emphasizing four fundamental processes — selection, drift, dispersal and mutation — that underlie microbial biogeographic patterns. We briefly introduce microbial biogeographic patterns and



**Figure 1 | The sliding scale of microbial taxonomic definitions and its influence on microbial biogeographic patterns.** **a** | Symbols at the tips of the phylogenetic tree represent individual genotypes. Taxonomic breadth is defined as the extent of taxa examined (for example, the genus *Pseudomonas* is a narrower breadth than all bacteria), and taxonomic resolution is the level of genetic variation at which taxa are defined. In this schematic, the differently coloured symbols represent four separate taxa at a 99% sequence similarity level. At a lower resolution (97% similarity, shown in black), only two taxa (represented by different shapes) are distinguished, whereas at a higher resolution all four taxa can be identified. **b, c** | Two possible biogeographic patterns; the outlined circles represent sampling locations that are separated by some spatial distance. Given the community in panel **a** and a 99% taxon definition, both endemism and a distance–decay pattern are observed in panel **b** (the yellow taxon is present only in the two circles on the left, and the blue taxon is present only in the rightmost circle). However, given a lower taxon resolution (97%), endemism would no longer be detectable (both shapes appear in all locations), but a distance–decay pattern would remain (more circles are seen on the left-hand side of the figure and more crosses on the right-hand side). By contrast, panel **c** depicts a case in which no taxa are endemic, regardless of the taxonomic resolution, but a distance–decay pattern exists at high taxonomic resolution (the abundances of the yellow and purple taxa decrease from left to right, whereas the abundances of the blue and orange taxa increase).

#### Taxon

A group into which related organisms are classified. For microorganisms, taxa are usually defined by sequence similarity of one or more genomic regions. This includes the possibility that taxa are defined by entirely unique genomes (at the highest genetic resolution possible).

#### Community

All individuals of a defined set of many taxa within an area (for instance, all bacterial taxa).

#### Taxonomic breadth

The extent of taxa considered.

#### Cosmopolitan

Having a widespread distribution, present almost everywhere.

discuss how the four processes interact to determine a commonly described pattern, the distance–decay relationship. We further show that the current published literature on microbial distance–decay patterns provides evidence for the interplay of these processes, even though the ability to detect a process can be conditional on the taxonomic, genetic and spatial resolution of the study. Last, we propose that a more direct consideration of these processes, including experimental tests, should drive the future of microbial biogeography research.

#### Microbial biogeographic patterns

Plant and animal biogeographers often examine the interplay of spatial and temporal patterns; for example, when tracking a taxon's spatial patterns through the fossil record or when describing the distribution of a taxon with a range that varies seasonally. Likewise, biogeographers are documenting striking temporal patterns in microbial diversity at the same location, primarily in marine ecosystems<sup>7–9</sup>. These studies show not only that microbial composition is highly temporally

variable within years but also that this variation exhibits a seasonal pattern that recurs across years. However, the vast majority of microbial biogeography studies focus solely on current spatial patterns, so we concentrate on these below.

**Evidence of biogeographic patterns.** Evidence that microorganisms display spatial biogeographic patterns falls into at least two categories. The first is the existence of endemic taxa and of nonrandom relationships in the similarity of taxa across a landscape. By definition, an endemic taxon is restricted to a particular location, region or habitat type and therefore is not distributed evenly across the Earth. For this reason, endemism is the clearest demonstration of microbial biogeography. Similarly to larger organisms, some microorganisms seem to be endemic to distinct geographic regions<sup>10,11</sup>. Other taxa are clearly endemic to specific habitat types: for example, those inhabiting extreme environments such as hot springs<sup>12–14</sup>. Finally, recent global surveys indicate that most bacteria are restricted to broad habitat types, as there is little overlap among bacterial taxa found in soils, sediments, freshwater and seawater<sup>15,16</sup>.

The second category of evidence is the detection of genetic similarity patterns among microbial taxa in different locations. One approach, adapted from population genetics, compares the genetic diversity within locations to that among locations, yielding a measure of divergence among locations<sup>13,17</sup>. Traditionally, these metrics have focused on a fine taxonomic resolution (often at the level of unique sequences) within a narrow taxonomic breadth, such as the sequence diversity within one 'species' (FIG. 1a; and as defined by REFS 18,19). Recently, however, these metrics have been applied to broader taxonomic breadths<sup>20,21</sup>: for example, to examine all of the diversity within a genus<sup>11,12,22</sup> or the entire bacterial community<sup>20,23</sup>. In each case, these studies find evidence that different locations harbour microorganisms that differ in genotypic composition.

A related approach seeks to correlate taxonomic similarity with geographic distance. A decline in similarity with increasing geographic distance (known as a distance–decay relationship), indicates not only that composition is different among locations but also that this variation is spatially autocorrelated (that is, it correlates with spatial distance)<sup>24–27</sup>. In population genetics, this pattern has long been studied at a relatively narrow taxonomic breadth and fine resolution to test for evidence of 'isolation by distance' or the tendency for populations of species that are close to one another to be more genetically similar to each other than to distant populations<sup>18,19</sup>. Distance–decay relationships have been repeatedly observed for microorganisms in a range of habitats at various taxonomic resolutions<sup>10,25,26,28–30</sup>.

**Limits to detecting patterns.** Not every study thus far has found evidence for microbial biogeographic patterns<sup>31,32</sup>, and some microbial taxa may indeed be cosmopolitan. However, there are several considerations to make before one discounts or embraces a microbial biogeographic pattern. First, the presence of cosmopolitan

microorganisms — the pigeons of the microbial world — does not exclude the possibility that other microbial taxa have endemic distributions (FIG. 1b). In fact, microbial studies commonly find evidence for both ubiquitous taxa and restricted taxa<sup>33</sup>. Moreover, microbial endemism is difficult to confirm because many rare taxa will not be detected in a sample even when they are present in the location, falsely suggesting a more restricted distribution than is actually the case<sup>9,34,35</sup>. Conversely, the presence of a taxon does not distinguish whether it is an active member of the community or a dormant transient (BOX 1).

Second, biogeographic patterns can exist in the absence of endemism. Many microbial diversity patterns, such as the distance–decay relationship, do not rely on the existence of endemic taxa (FIG. 1c). Even if every microbial taxon is everywhere, nonrandom spatial variation in the relative abundance of these taxa is a biogeographic pattern.

### Box 1 | What is dispersal limitation?

Many microbial biogeography studies focus primarily on the process of dispersal and, in particular, on the question of whether microorganisms are dispersal limited (that is, whether they show restriction in movement to and/or establishment at a location). Some of the debate about the existence of dispersal limitation stems from different definitions of dispersal. Here, we define dispersal as the movement and successful establishment of an individual (and the taxon it represents) from one location to another through passive or active mechanisms. Successful establishment means metabolic activity and at least some reproduction in the new location, rather than simply the presence of an individual or its taxon. This definition of dispersal is consistent with the concept of gene flow in population genetics, as only migrants that reproduce contribute to gene flow<sup>60</sup>. To draw an analogy with larger organisms, this distinction means that a plant community includes seedlings and adults but not the seed bank, and that an animal community does not include infrequent vagrants. For microorganisms, demonstrating establishment is not always feasible; therefore, detection is usually considered evidence of establishment, even though many microorganisms present in a location might not be metabolically or reproductively active<sup>9,34,93,108</sup>.

Thus, organisms can show dispersal limitation either if their movement to a new location is restricted or if establishment of individuals in a new location is hindered. Generally, restricted movement occurs when the probability of movement is uneven over space. For instance, an offspring or daughter cell that is more likely to disperse to a nearby location than to one further away (for example, if progeny cells remain spatially aggregated in the vicinity of the parent<sup>109</sup>) shows dispersal limitation, even if sometimes dispersal occurs everywhere.

Successful establishment depends on many factors, such as the presence of a suitable habitat and priority effects (when the prior occupation of a habitat precludes new individuals from colonizing that space)<sup>41,110</sup>. Priority effects may be neutral (whoever gets there first has an advantage regardless of taxon identity) or they may depend on a taxon's ability to compete with the particular taxa already present. Direct evidence that priority effects can influence microbial composition has been observed in laboratory microcosms<sup>111–113</sup>, but the importance of this mechanism in more natural habitats is unknown.

Finally, it is important to recognize that dispersal can depend on both neutral factors (which are independent of taxon identity) and deterministic factors. For instance, dispersal may be neutral when rates depend on population size, as individuals of abundant taxa should have a greater dispersal potential than individuals from low abundance populations, regardless of taxon identity<sup>41</sup>. However, microbial taxa can also vary in dispersal ability owing to deterministic traits such as spore formation, morphological features and habitat specificity. Establishment may also depend on deterministic factors such as physiological traits that influence habitat specificity or competitive ability (for example, autotrophs versus heterotrophs), habitat features (for example, availability of space and resources) and seasonal variation in favourable environmental conditions. Because both neutral and deterministic factors may contribute to dispersal, dispersal limitation by itself is not evidence that neutral processes are at work.

Finally, the ability to detect a biogeographic pattern, including endemism or a distance–decay relationship, may depend on taxonomic resolution. Specifically, some patterns may be more detectable at finer resolutions than at coarser resolutions<sup>2,10,36</sup> (FIG. 1b). Thus, the lack of a pattern using one taxonomic definition is not evidence that biogeographic patterns do not exist.

### The processes driving the patterns

Having established the existence of microbial biogeographic patterns, many researchers have begun drawing from multiple theoretical frameworks to evaluate the processes that generate and maintain the patterns observed<sup>3,4,6,37,38</sup>. One approach aims to partition the relative effects of current environmental factors versus the effects of historical ones on microbial biogeographic patterns<sup>6</sup>. Other frameworks draw upon the field of general ecology, using metacommunity theory<sup>39,40</sup> and the neutral theory of biodiversity<sup>41,42</sup> to interpret microbial spatial patterns<sup>3,38,43–47</sup>. These three seemingly disparate frameworks emphasize different subsets of the same four processes, which, for larger organisms, are usually considered separately, according to whether they operate on ecological or evolutionary levels<sup>48</sup>.

*Evolutionary versus ecological processes.* Mechanisms that contribute to the genetic composition and diversity within species are traditionally considered to be micro-evolutionary processes. Typically, species are defined according to the biological species concept: as a group of individuals that potentially interbreed and genetically recombine. Evolutionary biologists have long recognized four processes that contribute to diversity within such species: mutation, selection, gene flow and genetic drift<sup>49–53</sup>. Mutation creates non-directed change in nucleotide sequences. Natural selection drives the proliferation (or removal) of those mutations on the basis of the differential survival and reproduction of individuals carrying those mutations. Potential selective factors include all of the physical, chemical and biotic features of an organism's environment. Within-species diversity is further influenced by gene flow (the movement and successful establishment of genotypes from one population to another) and by genetic drift (changes in genotype frequencies owing to chance demographic events, such as births and deaths).

By contrast, mechanisms shaping the composition and diversity among species (and within and among multispecies communities) are classically referred to as ecological processes. Vellend<sup>48</sup> proposed that these processes can also be divided into four classes that are conceptually parallel to the four evolutionary processes: speciation, selection, dispersal and ecological drift. In this case, speciation adds new species diversity (as defined by the biological species concept). Selection alters the relative abundance of species on the basis of their ability to survive and reproduce. Selective factors among species are similar to those operating within species and include physical, chemical and biotic forces, such as the outcome of competition, predation and mutualism. Finally, species diversity within and among communities

is influenced by dispersal (the movement and successful establishment of a species to a new location) and by ecological drift (changes in the frequencies of species in a location owing to chance demographic fluctuations). Even for larger organisms, evolutionary and ecological processes are often so closely intertwined that their separation seems artificial<sup>54,55</sup>, although the distinction is particularly blurry for microorganisms. Currently, research on the drivers of microbial biogeographic patterns has primarily focused on what can arguably be called ecological processes because they examine the diversity and distribution among taxa, typically defined coarsely (for example, as >97% similarity of 16S or 18S ribosomal RNA genes). Although less prominent in the microbial biogeography literature, other studies have applied a population genetics perspective to investigate the evolutionary processes that drive diversity within taxa of relatively narrow breadth<sup>13,17,56–58</sup>. Despite this division, the boundary between ecological (among taxa) and evolutionary (within taxa) processes is as arbitrary and alterable as the microbial taxon definition itself.

#### *Four processes that shape microbial biogeography.*

The lack of a clear microbial species does not necessarily hinder the investigation of microbial biogeographic processes. Instead, we propose a focus on the same four fundamental processes — selection, drift, dispersal and mutation — without attempting to separate them into ecological and evolutionary levels. For instance, selection acts on many biological levels, including genes, genotypes, cells, individuals and taxa<sup>59,60</sup>; thus, it makes sense to consider selection generally for microbial taxa, which can be defined at many levels. Likewise, for practical purposes, microbial gene flow (the movement and establishment of genotypes) is measured in the same way as dispersal (the movement and establishment of taxa defined by genotypes). Even in the case of diploid, sexual microorganisms such as fungi, taxa are often characterized by haploid sequences rather than by the species concept.

Similarly, ecological and evolutionary drift cannot be distinguished for asexual organisms. Although some might argue that ecological drift occurs at a much faster rate than evolutionary drift, even this distinction falls apart because of the flexible nature of the microbial taxon. Consider an extreme case, in which bacterial taxa are defined by unique nucleotide sequences (FIG. 1a). Imagine then that stochastic changes occur in relative abundance among all such bacterial taxa in a community. Because the changes are among taxa, one could argue that ecological drift has altered bacterial diversity in this community. Now consider only the changes that happened during this same time within one narrow lineage (for instance, just within all *Escherichia coli* strains). Clearly, these same changes are not due to a different process (evolutionary drift) simply because the taxonomic focus (all bacteria versus only *E. coli*) has changed.

A less obvious parallel between processes is the case of mutation and speciation. However, mutation, in conjunction with the other three processes mentioned

above, is ultimately responsible for diversification among taxa (including speciation). Finally, one might argue that the four processes are not exclusive. For instance, previously we considered extinction a separate process<sup>6</sup>, but now we classify it as the result of drift or selection<sup>48</sup>.

In sum, under this proposed merged framework, just four processes act at all taxonomic scales (resolution and breadth) to create and maintain microbial biogeographic patterns. Of course, there can be mechanistic differences in how these processes operate within and across various taxonomic scales. Most notably, selection within sexual populations differs mechanistically from selection acting on taxa as defined by a genetic sequence. Similarly, accounting for reproductive isolation mechanisms in the speciation process is more complex than accounting for the basic mutation process. Thus, it remains to be tested at which taxonomic scales these differences emerge for microorganisms. However, for the purposes of understanding microbial biogeography, we propose that more insight will be gained by considering how these processes influence diversity across taxonomic scales, rather than by dividing them into ecological versus evolutionary categories.

A key reason that such a framework is possible is that most studies investigate the biogeography of microbial genes, not that of entire organisms or genomes (but see REFS 45,61). These marker genes are chosen in large part because they are highly conserved, they are rarely subject to horizontal gene transfer (HGT) and they maintain robust phylogenetic relationships<sup>62</sup>. Thus, recombination and HGT should be of minor importance to the patterns observed, although an interesting avenue of future research lies in comparing biogeographic patterns of conserved marker genes with those of entire genomes. Genome-wide approaches may additionally offer more robust species classifications<sup>63–65</sup>, providing the opportunity to test whether species definitions (reviewed in REF. 66) better explain microbial biogeographic patterns than operational taxonomic definitions.

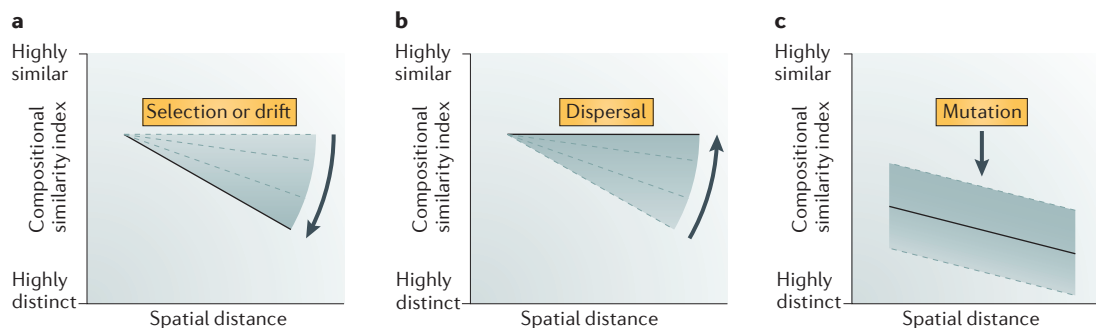
#### **Process to pattern: distance–decay**

To demonstrate how the processes of selection, drift, dispersal and mutation shape biogeographic patterns, we first abstractly consider their influence on the most commonly studied pattern in the published microbial literature, the distance–decay relationship. Overall, selection and drift generate the relationship (a negative correlation between compositional similarity and geographic distance), dispersal counteracts it, and mutation modifies its variance (the spread of data points) and height (FIG. 2).

The process of selection generally differentiates microbial composition among locations. At any one location, the environment selects for taxa (whether defined by individual genotypes or OTUs) that are relatively better adapted to the local conditions, as long as taxa vary in their response to those conditions<sup>48,52</sup>. At the same time, selective factors are often organized spatially, as in a gradient, and are thus spatially autocorrelated. In such a case, selection will tend to produce a distance–decay relationship, in which compositional

Horizontal gene transfer  
Transfer of genetic material  
between independent  
organisms other than transfer  
by direct descent.





**Figure 2 | The effect of the four processes on the relationship between compositional similarity and spatial distance.** **a** | Selection and drift increase the strength of the distance–decay relationship (they steepen the slope). **b** | Dispersal weakens the distance–decay relationship (it flattens the slope). **c** | Mutation decreases the similarity between locations, regardless of the distance between them.

similarity between any two locations decreases as the geographic distance between them increases (FIG. 2a). For instance, in a coastal estuary, salinity is highly spatially autocorrelated, with highly saline conditions near the ocean and fresher conditions further inland. Microbial taxa are also highly sensitive to salinity; therefore, two locations near the ocean are likely to be more similar to one another in microbial composition than to a location further inland<sup>67</sup>.

Similarly to selection, drift differentiates microbial composition over space. Chance events, including stochastic differences among taxa in births, deaths and migration, differ among locations, regardless of the surrounding environment or phenotypic variation among taxa. Models derived from neutral theory, which does not take selection into account, demonstrate that drift, even in the absence of selection, can create a distance–decay relationship<sup>18,68,69</sup>. As with selection, drift strengthens the distance–decay relationship (it steepens the slope) (FIG. 2a). Importantly, drift must interact with dispersal to create a distance–decay pattern<sup>41</sup>: without dispersal, drift would create a patchy distribution of microorganisms and, hence, variation in composition that is not spatially autocorrelated<sup>69</sup>. By contrast, with some dispersal limitation (BOX 1), microbial composition would be more similar between nearby locations than between those further apart, as chance events at one location would influence nearby composition. Thus, one might expect that the importance of drift for the distance–decay pattern will be greater for microorganisms that are subject to relatively restricted dispersal, perhaps at large spatial scales (for example, between continents) or in less fluid environments (for example, subsurface habitats).

By contrast, as dispersal rates increase, local composition increasingly reflects that of the newly established colonizers, rather than the result of subsequent selection by underlying environmental conditions and/or local chance events<sup>18,40,49,69</sup>. At some point, dispersal will entirely counteract compositional differentiation imposed by drift and/or selection and eliminate the distance–decay relationship (that is, flatten the slope of the distance–decay curve) (FIG. 2b). Thus, all else being equal, the distance–decay relationship should be relatively weak in habitats where dispersal is high,

such as in the pelagic marine environment, where ocean currents facilitate microbial dispersal<sup>30</sup>. By contrast, the distance–decay relationship should be stronger when dispersal is more limited, such as across disconnected freshwater bodies<sup>44</sup>.

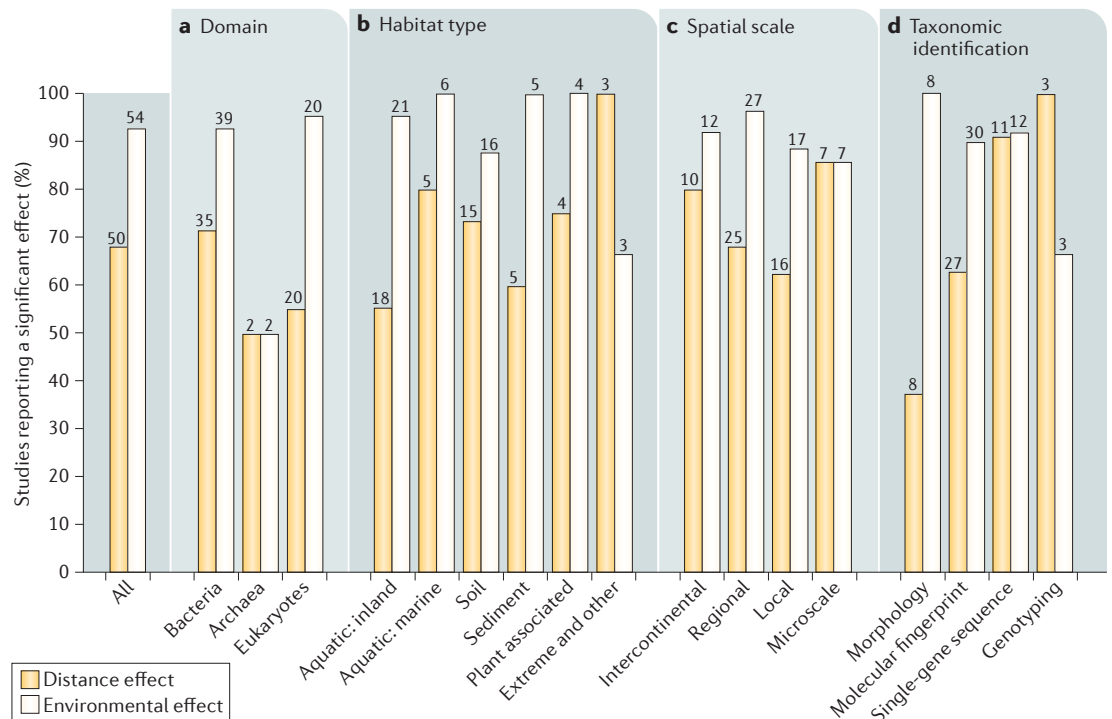
Last, mutation modifies the distance–decay relationship by increasing local genetic diversity across all locations. Because the same mutation rarely occurs twice, this process essentially adds ‘noise’ to the genetic diversity present among locations<sup>18,70</sup>. Thus, mutation makes composition more distinct between locations, lowering the height (not the slope) of the distance–decay curve and increasing its variance<sup>18,69,70</sup> (FIG. 2c). For studies that focus on coarsely defined taxa, mutation is unlikely to affect discernible biogeographic patterns, as new mutations create within-taxon variation that is not taken into account. However, if a variable genetic region is considered at a fine taxonomic resolution (for example, the internal transcribed spacer (ITS) region at 99% similarity), then mutation might noticeably influence the average compositional similarity among locations and thus the distance–decay relationship. Although we are not aware of any distance–decay studies that examine the effect of mutation directly, given the short generation times and fast growth rates of many microorganisms, mutation might indeed add detectable variability in microbial composition at fine resolutions.

### Pattern to process: a literature review

The challenge remains to identify the relative importance of the four processes on observed biogeographic patterns. We conducted a non-exhaustive review of studies that attempted to disentangle the relative effects of contemporary selection versus historical processes on the distance–decay relationship (see legend of FIG. 3 for details). Generally, the studies took a two-step approach. First, they tested for a correlation between microbial composition and measured environmental variables across many sampled locations. This correlation is evidence for the influence of contemporary selection; that is, the influence of the current environment on the current distribution of microbial diversity. After controlling for this environment effect, they then tested whether geographic

#### Environment effect

A correlation of biotic composition with measured environmental variables after controlling for the influence of geographic distance.



**Figure 3 | Studies reporting a significant correlation between microbial composition and spatial distance or an environmental variable. a–d** | Our literature review ( $n = 54$ ) included studies of bacteria, archaea and/or microbial eukaryotes, but excluded those that focused on known pathogens, viruses or animal-associated microorganisms, the distributions of which may be driven largely by those of their hosts. All studies used statistical methods to test for the effect of horizontal spatial distance or environment while controlling for the effect of the other factor (for example, redundancy analysis with variation partitioning<sup>76,77</sup> or partial Mantel tests<sup>114</sup>). In most cases, the spatial parameter reported was geographic distance, but in several cases other proxies for dispersal likelihood were used (for example, habitat connectivity or lake-water retention time). Sample size (the number of studies that statistically tested for the effect in each category) is shown above each bar. In panel **c**, scale definitions are as follows: intercontinental, >5,000 km; regional, 101–5,000 km; local, 0–100 km; microscale, 0–1 km; note that microscale is a subset of the local scale. Studies were categorized according to the maximum scale used in the distance test. Several studies divided their data sets into different spatial scales and tested for an effect at each scale separately, and are thus included in each relevant spatial category. In panel **d**, the morphology category includes standard morphological species identification techniques. The molecular fingerprint category includes ARISA (automated ribosomal intergenic spacer analysis), DGGE (denaturing gradient gel electrophoresis), RFLP (restriction fragment length polymorphism), TRFLP (terminal RFLP) and single-strand polymorphism techniques. The single-gene sequence category refers to traditional and pyrosequencing approaches targeted at a single genetic region. Genotyping, the most sensitive characterization category, includes whole-genome and multilocus techniques, such as ERIC-PCR (enterobacterial repetitive intergenic consensus PCR) and multilocus sequencing. Studies based on PLFA (phospholipid fatty acid) analysis ( $n = 4$ ) and qPCR (quantitative PCR;  $n = 1$ ) were excluded from the analysis in panel **d**.

distance explained any of the remaining variation in microbial composition. Such a distance effect indicates that processes other than current selection affect spatial variation in microbial composition<sup>6,27,71,72</sup>. In this case, history (encompassing the very recent to the geological past) has left a legacy on present-day microbial composition, either through past selection or drift. Importantly, at least some dispersal limitation (BOX 1) is required for past events to leave a contemporary signature, otherwise differentiation owing to past selection or drift would be counteracted (FIG. 2). As a result, a significant distance effect can be interpreted as evidence of dispersal limitation. Collectively, the processes behind a distance effect — drift and/or past selection along with dispersal limitation — are often referred to as historical processes.

The literature survey allowed us to address the following two questions. What is the relative importance of contemporary environmental selection and/or historical processes to observed patterns? And does this importance vary by taxon, habitat, spatial scale or taxonomic resolution of the study?

**Evidence for contemporary selection and historical processes.** Our survey indicates that both contemporary selection and historical processes shape microbial biogeographic patterns. Most of the studies (92.6%) found a significant correlation between microbial composition and at least one measured environmental or habitat feature, demonstrating that selection imposed by the contemporary environment has a prominent role in shaping microbial biogeographic patterns (FIG. 3).

#### Distance effect

A correlation of biotic composition with geographic distance after controlling for the influence of the contemporary environment.

Only four out of the 54 studies reviewed observed no correlation between composition and environmental variables<sup>13,73–75</sup>. Most of the studies (68%) also reported a significant distance effect, providing evidence that historical processes, including dispersal limitation, influence microbial composition (FIG. 3).

Among the studies examined, contemporary selection had a greater effect on microbial composition than historical processes. A study was more likely to find a significant effect caused by environmental variables than one caused by geographic distance (FIG. 3), and more variation in microbial composition could be explained by environmental variables (26.9%) than by geographic distance (10.3%;  $t(36) = 3.27$ ,  $P = 0.0053$ ; FIG. 4). When the unique and combined effects of environment and distance were totalled, on average the studies explained 49.7% of the total variation in community composition (FIG. 4). This result is remarkably similar to that of a similar literature review on the biogeography of larger organisms<sup>76</sup>, in which 22% of compositional variation could be explained by the environment alone, 16% by geographic distance alone and 48% by a combination of the two. Thus, contemporary selection seems to be more important than historical processes to the biogeography of both small and large organisms.

A caveat to interpreting the distance effect, and therefore the importance of dispersal limitation, is the problem of unmeasured environmental variables. A spurious distance effect will result if any spatially autocorrelated selective factors are not accounted for by the measured environmental variables<sup>27,77</sup>. In fact, a complete quantification of all selective variables is practically impossible to achieve, so the distance effect is probably almost always overestimated. For example, biotic variables, such as the abundance or composition of other organisms present, are rarely taken into account. However, when they are considered, they often correlate with microbial composition<sup>78–80</sup>, suggesting that biotic interactions are an often disregarded selective force capable of influencing microbial distributions.

#### **When and where are historical processes important?**

The relative importance of contemporary selection versus historical processes probably depends on a taxon's physiological traits. However, studies focusing on eukaryotic microorganisms were just as likely to find a distance effect as those on bacteria and archaea (FIG. 3a;  $\chi^2(2, n = 57) = 1.71$ ,  $P = 0.43$ ), even though dispersal rates are expected to increase with decreasing body or cell size<sup>81</sup>. Of course, the lack of a domain-level pattern does not exclude the possibility that important trait variation occurs among finer taxa. For instance, bacterial taxa that produce spores or cysts might have greater dispersal capabilities than those without such traits<sup>82</sup>.

Habitat type might also affect the importance of historical processes, as highly connected habitats and aquatic substrates should enable more dispersal than isolated habitats and solid substrates<sup>44,83,84</sup>. However, the proportion of studies detecting a distance effect was also similar across habitat types (FIG. 3b;

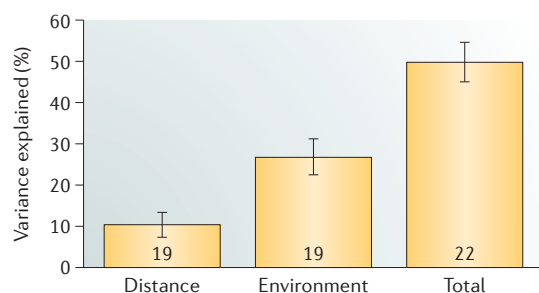
$\chi^2(5, n = 50) = 3.46$ ,  $P = 0.63$ ). Furthermore, distance effects were detected in the ocean as often as in soils or inland aquatic environments, although there was a trend for inland aquatic environments to display distance effects less often than soils.

Historical processes might also be more evident at larger geographic scales, owing to decreasing migration rates. Once again, however, the studies reviewed did not reveal a clear trend. The likelihood of detecting a distance effect did not vary by geographic scale (FIG. 3c;  $\chi^2(3, n = 58) = 1.79$ ,  $P = 0.63$ ). If anything, a distance effect was most often observed at small (0–1 km) or very large, intercontinental (>5,000 km) scales. This result is highlighted by several studies that directly compared multiple spatial scales. For example, dispersal-related factors in diatoms were found to increase in relative importance with increasing maximum distance between lakes<sup>85</sup>. Similarly, a distance effect was apparent at only the largest spatial scales for freshwater bacteria in two other studies<sup>29,47</sup>. By contrast, community composition was related to distance at only the smallest scale (<5 km) for salt marsh bacteria, not at the regional or global scale<sup>86</sup>. Such a small-scale distance effect might be due to microbial aggregation<sup>87</sup>, which can be caused by dispersal limitation (BOX 1).

Finally, the ability to detect historical processes might depend on taxonomic resolution, as higher resolutions take into account more compositional variation<sup>4,26,38</sup>. Indeed, the likelihood of detecting a distance effect increased with increasing resolution (FIG. 3d;  $\chi^2(3, n = 47) = 7.68$ ,  $P = 0.053$ ). Studies using whole-genome or multilocus genotyping techniques and full-sequence comparisons were more likely to detect a distance effect than studies using single-gene fingerprinting techniques or morphological techniques. In addition, among those studies that sequenced the 16S gene, pyrosequencing studies (which usually consider a lower taxonomic resolution than those using cloning and Sanger sequencing<sup>88</sup>) were less likely to detect an effect (43% of the time) than cloning studies (80%). This trend is consistent with several studies that directly compared different levels of taxonomic resolution<sup>26,85,89,90</sup>. For example, one study<sup>10</sup> found stronger evidence for endemism in soil *Pseudomonas* spp. with a high-resolution genotyping method than with fingerprinting methods<sup>10</sup>. In marine *Prochlorococcus* spp., community composition correlated with dispersal rate at only the highest taxonomic resolution examined<sup>90</sup>.

#### **Moving forward**

The literature analysis suggests that both contemporary selection and historical processes (including dispersal limitation) shape the distribution of at least some microorganisms from all domains of life and a range of habitat types, spatial scales and taxonomic resolutions (FIG. 3). However, selection seems to have a stronger influence than historical processes (FIGS 3,4), as has been observed for larger organisms<sup>76</sup>. Beyond these broad conclusions it remains difficult to disentangle the relative importance of selection, drift, dispersal and mutation by analysing distance–decay patterns<sup>41,42,91</sup>.



**Figure 4 | Explained variance in microbial composition.** The percentage of variance explained by the unique effect of horizontal geographic distance, the unique effect of measured environmental variables and the sum of the unique and combined fractions in studies that carried out a redundancy analysis (RDA) with variance partitioning<sup>76,77,114</sup>. The total number of studies that reported RDA values (sample size) is shown within each bar, and standard error is represented by error bars. The mean variance explained by the unique effects of the environment is significantly greater than the mean variance explained by the unique effects of geographic distance ( $t(36) = 3.27, P = 0.0053$ ).

Below, we offer suggestions for how the field might advance by emphasizing biogeographic processes over patterns.

**Focusing the question.** Although most studies focus much of their interpretation on dispersal, a distance effect is created by a lack of dispersal (dispersal limitation) and the resulting process of drift. Thus, one might ask instead: what is needed to detect drift in microbial communities? When considering this question, it is important to note that drift has a larger impact when dispersal is low and populations are small<sup>41,52</sup>. As a result, drift is often assumed to be irrelevant for microorganisms owing to their high dispersal potential and large population sizes. This assumption may not be true for several reasons. First, even when dispersal rates are high, models suggest that drift can create a strong distance–decay pattern<sup>41</sup>. Second, although the total number of all microbial individuals in a community may be very large, drift might still act on rare taxa<sup>92</sup>; indeed, most microbial taxa within a community seem to be rare<sup>34</sup>. In addition, many individuals may not be actively reproducing (that is, they may be dormant<sup>93</sup>) or they may be divided into subpopulations<sup>63</sup>, thereby reducing local ‘effective’ population sizes.

Given the conditions leading to drift, sampling methods that target rare and/or active microbial populations will be more likely to detect drift. Similarly, because selection seems to be a nearly universally important process in shaping microbial biogeography, a study designed to investigate drift should minimize environmental variation (and therefore selection). By contrast, sampling across environmental gradients or habitat types (as most studies to date have done) emphasizes the strength of environmental selection, and therefore any effect of drift on microbial composition may go undetected.

Finally, as noted several times above, the ability to detect a biogeographic process can depend on taxonomic

resolution. However, it is important to clarify that it is not taxonomic resolution per se that matters, but the amount of genetic variation captured by this resolution. Specifically, for the detection of drift, the amount of selectively neutral variation (that is, genetic variation that has no effect on fitness) of the taxonomic marker is important. If selection at the marker locus is high and/or there are few neutral sites (genetic sites at which variation has no effect on fitness) within the marker, then differentiation by drift may not be recognizable at that marker<sup>18,94</sup>. This suggests that commonly used markers, such as the 16S rRNA gene, may not be ideal candidates for the detection of drift if they encompass little neutral variation. Indeed, if selection is operating on a genetic marker or a morphological trait used to define taxa, a distance effect may not be observed because there is not enough non-selective variation to detect drift. Similarly, the use of multiple loci can improve the detection of drift by increasing the amount of neutral variation covered. For instance, one study<sup>13</sup> detected a distance effect in *Sulfolobus* spp. populations at smaller spatial scales only when analysing multiple genetic loci.

**Beyond patterns.** Part of the difficulty in inferring the driving processes from biogeographic patterns is that the processes occur along a continuum of space and time<sup>95</sup>, whereas discretely sampled patterns represent an integration of these processes over time. To address the effects of these processes over time, both experiments of microbial community assembly and temporal studies are needed. There are few examples of experimental studies, but those that do exist provide direct evidence that historical processes influence microbial community composition, whether through the legacy of past environmental conditions<sup>96</sup> or through dispersal limitation<sup>97–99</sup>. For instance, one study<sup>99</sup> identified a specific temporal window during which the effects of dispersal limitation were apparent before other processes eclipsed its impact. Thus, the relative importance of a process may depend on when during assembly the community is examined. Furthermore, recent temporal survey studies have highlighted the importance of seasonality as a driver of microbial distribution and activity<sup>7–9</sup>, which has implications for the interpretation of spatial patterns observed at one point in time.

An additional approach is to directly address the propensity of particular microbial taxa to be affected by any of the four processes discussed here. For example, dispersal potential can be estimated by characterizing microbial composition in dispersal vectors such as air, dust and rain<sup>100–105</sup>. Measurements of phenotypic traits that facilitate dispersal can also be valuable. One study<sup>82</sup> found that even coarse knowledge about dispersal-related traits from relatively few cultured taxa could explain the observed biogeographic patterns. Finally, new mathematical models, perhaps adapted from population genetics and ecological theory, could shed light on biogeographic processes. Such models might, for example, explore the potential rates of one process, relative to another, that are required to produce an observed distance–decay relationship.



## Conclusion

In the past decade the field of microbial biogeography has made significant strides towards demonstrating that microorganisms display biogeographic patterns. The biogeography of microorganisms, like that of all organisms, is undoubtedly governed by the evolutionary and ecological interplay of four major processes: selection, drift, dispersal and mutation. However, when, where, why and how much each process contributes to these patterns remains unresolved for microorganisms, as it does for all

organisms<sup>106</sup>. At the same time, microbial biogeography is close on the intellectual heels of the broader field of biogeography, and the ease of carrying out multi-generational experiments with microorganisms offers the possibility to test hypotheses that are not possible in larger organisms<sup>107</sup>. With increasingly powerful survey tools, creative experiments, temporal data sets and new theoretical models, the next era of microbial biogeography promises to transform our understanding of the processes shaping all biodiversity.

- Lomolino, M. V., Riddle, B. R., Whittaker, R. & Brown, J. H. *Biogeography* (Sinauer Associates, 2010).
- Green, J. & Bohannan, B. J. M. Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21**, 501–507 (2006).
- Lindström, E. S. & Langenheder, S. Local and regional factors influencing bacterial community assembly. *Environ. Microbiol. Rep.* **4**, 1–9 (2012).
- Ramette, A. & Tiedje, J. M. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb. Ecol.* **53**, 197–207 (2007).
- Foissner, W. Biogeography and dispersal of microorganisms: a review emphasizing protists. *Acta Protozool.* **45**, 111–136 (2006).
- Martiny, J. B. H. et al. Microbial biogeography: putting microorganisms on the map. *Nature Rev. Microbiol.* **4**, 102–112 (2006).  
**This article outlines the environment versus the history approach to assessing the processes underlying microbial biogeographic patterns; this Review serves as a conceptual update to this framework.**
- Fuhrman, J. A. et al. Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl Acad. Sci. USA* **103**, 13104–13109 (2006).
- Gilbert, J. A. et al. Defining seasonal marine microbial community dynamics. *ISME J.* **6**, 298–308 (2012).
- Caporaso, J. G., Paszkiewicz, K., Field, D., Knight, R. & Gilbert, J. A. The Western English Channel contains a persistent microbial seed bank. *ISME J.* 10 Nov 2011 (doi:10.1038/ismej.2011.162).  
**A recent example of temporal variation in microbial distributions, highlighting the issue of rarity and dormancy in defining microbial community composition.**
- Cho, J. C. & Tiedje, J. M. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66**, 5448–5456 (2000).
- Oakley, B. B., Carbonero, F., van der Gast, C. J., Hawkins, R. J. & Purdy, K. J. Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations. *ISME J.* **4**, 488–497 (2010).
- Takacs-Vesbach, C., Mitchell, K., Jackson-Weaver, O. & Reysenbach, A. L. Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. *Environ. Microbiol.* **10**, 1681–1689 (2008).
- Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**, 976–978 (2003).  
**This classic study in microbial biogeography finds evidence for both endemism and distance–decay and uses a population genetics approach to interpret those patterns.**
- Papke, R. T., Ramsing, N. B., Bateson, M. M. & Ward, D. M. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**, 650–659 (2003).
- Nemergut, D. R. et al. Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* **13**, 135–144 (2011).
- Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl Acad. Sci. USA* **104**, 11436–11440 (2007).
- Vos, M. & Velicer, G. J. Isolation by distance in the spore-forming soil bacterium *Mycobacterium xanthus*. *Curr. Biol.* **18**, 386–391 (2008).
- Slatkin, M. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264–279 (1993).
- Wright, S. The genetical structure of populations. *Ann. Eugen.* **15**, 323–354 (1951).
- Martin, A. P. Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl. Environ. Microbiol.* **68**, 3673–3682 (2002).
- Hu, X.-S., He, F. & Hubbell, S. P. Community differentiation on landscapes: drift, migration and speciation. *Oikos* **118**, 1515–1523 (2009).
- Bahl, J. et al. Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nature Commun.* **2**, 163 (2011).
- O'Mullan, G. D. & Ward, B. B. Relationship of temporal and spatial variabilities of ammonia-oxidizing bacteria to nitrification rates in Monterey Bay, California. *Appl. Environ. Microbiol.* **71**, 697–705 (2005).
- Soininen, J., McDonald, R. & Hillebrand, H. The distance decay of similarity in ecological communities. *Ecography* **30**, 3–12 (2007).
- Green, J. L. et al. Spatial scaling of microbial eukaryote diversity. *Nature* **432**, 747–750 (2004).
- Horner-Devine, M. C., Lage, M., Hughes, J. B. & Bohannan, B. J. M. A taxa-area relationship for bacteria. *Nature* **432**, 750–753 (2004).
- Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* **26**, 867–878 (1999).
- Castelley, G. et al. Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proc. Natl Acad. Sci. USA* **107**, 12952–12957 (2010).
- Soininen, J., Korhonen, J. J., Karhu, J. & Vetterli, A. Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. *Limnol. Oceanogr.* **56**, 508–520 (2011).
- Hewson, L., Steele, J. A., Capone, D. G. & Fuhrman, J. A. Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. *Mar. Ecol. Prog. Ser.* **311**, 67–77 (2006).
- Queloz, V., Sieber, T. N., Holdenrieder, O., McDonald, B. A. & Grünig, C. R. No biogeographical pattern for a root-associated fungal species complex. *Global Ecol. Biogeogr.* **20**, 160–169 (2011).
- Finlay, B. J. Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063 (2002).
- Pommier, T. et al. Global patterns of diversity and community structure in marine bacterioplankton. *Mol. Ecol.* **16**, 867–880 (2007).
- Sogin, M. L. et al. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl Acad. Sci. USA* **103**, 12115–12120 (2006).
- Curtis, T. P., Sloan, W. T. & Scannell, J. W. Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci. USA* **99**, 10494–10499 (2002).
- Fulthorpe, R. R., Roesch, L. F. W., Riva, A. & Triplett, E. W. Distantly sampled soils carry few species in common. *ISME J.* **2**, 901–910 (2008).
- Hedlund, B. P. & Staley, J. T. in *Microbial Diversity and Bioprospecting* (ed. Bull, A. T.) 225–231 (ASM Press, 2003).
- Logue, J. B. & Lindstrom, E. S. Biogeography of bacterioplankton in inland waters. *Freshwater Rev.* **1**, 99–114 (2008).
- Holyoak, M., Leibold, M. A. & Holt, R. D. (eds) *Metacommunities: Spatial Dynamics and Ecological Communities* (Univ. of Chicago Press, 2005).
- Leibold, M. A. et al. The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* **7**, 601–613 (2004).  
**A concise, conceptual summary of metacommunity theory.**
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).  
**The theoretical derivation of neutral theory as applied to multispecies communities.**
- Bell, G. Ecology — neutral macroecology. *Science* **293**, 2413–2418 (2001).
- Drakare, S. & Liess, A. Local factors control the community composition of cyanobacteria in lakes while heterotrophic bacteria follow a neutral model. *Freshwater Biol.* **55**, 2447–2457 (2010).
- Barberán, A. & Casamayor, E. O. Global phylogenetic community structure and  $\beta$ -diversity patterns in surface bacterioplankton metacommunities. *Aquat. Microb. Ecol.* **59**, 1–10 (2010).
- Keymer, D. P., Lam, L. H. & Boehm, A. B. Biogeographic patterns in genomic diversity among a large collection of *Vibrio cholerae* isolates. *Appl. Environ. Microbiol.* **75**, 1658–1666 (2009).
- Sloan, W. T. et al. Neutral assembly of bacterial communities. *FEMS Microbiol. Ecol.* **62**, 171–180 (2007).
- Van der Gucht, K. et al. The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proc. Natl Acad. Sci. USA* **104**, 20404–20409 (2007).
- Vellend, M. Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).  
**An introduction to the many theoretical frameworks in community ecology that could be used to explain species distributions and arguments for how all of these frameworks can be merged by focusing on just four fundamental processes.**
- Slatkin, M. Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792 (1987).  
**A general, conceptual description of the effects of mutation, drift, gene flow and selection on spatial variation in within-species diversity.**
- Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
- Dobzhansky, T. G. *Genetics and the Origin of Species* (Columbia Univ. Press, 1937).
- Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer Associates, 2007).  
**A general textbook reference for the evolutionary processes that generate diversity within species.**
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, 1983).
- Antonovics, J. The input from population genetics: “the new ecological genetics”. *Syst. Bot.* **1**, 233–245 (1976).
- Lenski, R. E. in *Ecology: Achievement and Challenge* (eds Press, M. C., Huntly, N. J. & Levin, S.) 25–45 (Blackwell Science, 2001).
- Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. M. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature* **418**, 171–174 (2002).
- Lenski, R. E. et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247 (2009).
- Travisano, M., Mongold, J. A., Bennett, A. F. & Lenski, R. E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995).
- Lewontin, R. C. The units of selection. *Annu. Rev. Ecol. Syst.* **1**, 1–18 (1970).
- Futuyma, D. J. *Evolution* (Sinauer Associates, 2005).
- Ramette, A. & Tiedje, J. M. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc. Natl Acad. Sci. USA* **104**, 2761–2766 (2007).
- Daubin, V., Moran, N. A. & Ochman, H. Phylogenetics and the cohesion of bacterial genomes. *Science* **301**, 829–832 (2003).
- Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**, 741–746 (2009).

64. Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl Acad. Sci. USA* **102**, 2567–2572 (2005).
65. Caro-Quintero, A. & Konstantinidis, K. T. Bacterial species may exist, metagenomics reveal. *Environ. Microbiol.* **14**, 347–355 (2011).
66. Achtman, M. & Wagner, M. Microbial diversity and the genetic nature of microbial species. *Nature Rev. Microbiol.* **6**, 431–440 (2008).
67. Hewson, I. & Fuhrman, J. A. Richness and diversity of bacterioplankton species along an estuarine gradient in Moreton Bay, Australia. *Appl. Environ. Microbiol.* **70**, 3425–3433 (2004).
68. Condit, R. *et al.* Beta-diversity in tropical forest trees. *Science* **295**, 666–669 (2002).
69. Hutchison, D. W. & Templeton, A. R. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**, 1898–1914 (1999).
70. Hedrick, P. W. Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313–318 (1999).
71. Ricklefs, R. E. A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.* **7**, 1–15 (2004).
72. Svenning, J. C. & Skov, F. The relative roles of environment and history as controls of tree species composition and richness in Europe. *J. Biogeogr.* **32**, 1019–1035 (2005).
73. Yergeau, E., Newsham, K. K., Pearce, D. A. & Kowalchuk, G. A. Patterns of bacterial diversity across a range of Antarctic terrestrial habitats. *Environ. Microbiol.* **9**, 2670–2682 (2007).
74. Casamayor, E. O., Reche, I., Pulido-Villena, E. & Morales-Baquero, R. Does ecosystem size determine aquatic bacterial richness? *Ecology* **86**, 1715–1722 (2005).
75. Baker, K. L. *et al.* Environmental and spatial characterisation of bacterial community composition in soil to inform sampling strategies. *Soil Biol. Biochem.* **41**, 2292–2298 (2009).
76. Cottenie, K. Integrating environmental and spatial processes in ecological community dynamics. *Ecol. Lett.* **8**, 1175–1182 (2005).
- A literature review that uses a metacommunity perspective to evaluate the relative importance of selection versus dispersal-driven processes for driving variation in the spatial distribution of primarily non-microbial organisms.**
77. Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055 (1992).
78. St-Arnaud, M. *et al.* Patterns of *Fusarium* community structure and abundance in relation to spatial, abiotic and biotic factors in soil. *FEMS Microbiol. Ecol.* **71**, 34–42 (2010).
79. Yergeau, E. *et al.* Influences of space, soil, nematodes and plants on microbial community composition of chalk grassland soils. *Environ. Microbiol.* **12**, 2096–2106 (2010).
80. Zinger, L. *et al.* Contrasting diversity patterns of crenarchaeal, bacterial and fungal soil communities in an alpine landscape. *PLoS ONE* **6**, e19950 (2011).
81. Fenchel, T. There are more small than large species. *Oikos* **68**, 375–378 (1993).
82. Bissett, A., Richardson, A. E., Baker, G., Wakelin, S. & Thrall, P. H. Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol. Ecol.* **19**, 4315–4327 (2010).
83. Cermeño, P. & Falkowski, P. G. Controls on diatom biogeography in the ocean. *Science* **325**, 1539–1541 (2009).
84. Chase, J. N. & Forbes, A. E. The role of habitat connectivity and landscape geometry in experimental zooplankton metacommunities. *Oikos* **96**, 433–440 (2002).
85. Verleyen, E. *et al.* The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos* **118**, 1239–1249 (2009).
86. Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D. & Horner-Devine, M. C. Drivers of bacterial  $\beta$ -diversity depend on spatial scale. *Proc. Natl Acad. Sci. USA* **108**, 7850–7854 (2011).
87. Morlon, H. *et al.* A general framework for the distance–decay of similarity in ecological communities. *Ecol. Lett.* **11**, 904–917 (2008).
88. Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A. & Versalovic, J. Metagenomic pyrosequencing and microbial identification. *Clin. Chem.* **55**, 856–866 (2009).
89. Harder, J., Schauer, R., Bienhold, C. & Ramette, A. Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. *ISME J.* **4**, 159–170 (2010).
90. Martiny, A. C., Tai, A. P. K., Veneziano, D., Primeau, F. & Chisholm, S. W. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ. Microbiol.* **11**, 823–832 (2009).
91. Cale, W. G., Henebry, G. M. & Yeakley, J. A. Inferring process from pattern in natural communities. *Bioscience* **39**, 600–605 (1989).
92. Hu, X. S., He, F. L. & Hubbell, S. P. Neutral theory in macroecology and population genetics. *Oikos* **113**, 548–556 (2006).
- A concise, conceptual summary and comparison of neutral theory as it applies to population genetics and ecology.**
93. Lennon, J. T. & Jones, S. E. Dormancy contributes to the maintenance of microbial diversity. *Proc. Natl Acad. Sci. USA* **107**, 5881–5886 (2010).
94. Karl, S. A. & Avise, J. C. Balancing selection at allozyme loci in oysters — implications from nuclear RFLPs. *Science* **256**, 100–102 (1992).
95. Ricklefs, R. E. History and diversity: explorations at the intersection of ecology and evolution. *Am. Nat.* **170**, S56–S70 (2007).
96. Keiser, A. D., Strickland, M. S., Fierer, N. & Bradford, M. A. The effect of resource history on the functioning of soil microbial communities is maintained across time. *Biogeosciences* **8**, 1477–1486 (2011).
97. Langenheder, S. & Szekely, A. J. Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISME J.* **5**, 1086–1094 (2011).
98. Ayarza, J. M. & Erijman, L. Balance of neutral and deterministic components in the dynamics of activated sludge floc assembly. *Microb. Ecol.* **61**, 486–495 (2011).
99. Bell, T. Experimental tests of the bacterial distance–decay relationship. *ISME J.* **4**, 1357–1365 (2010).
- An example of a manipulative experimental study to assess the mechanisms contributing to spatial variation in microbial community composition.**
100. Sharma, N. K. & Singh, S. Differential aerosolization of algal and cyanobacterial particles in the atmosphere. *Indian J. Microbiol.* **50**, 468–473 (2010).
101. Figuerola, J. & Green, A. J. Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. *Freshwater Biol.* **47**, 483–494 (2002).
102. McMahon, K. D., Jones, S. E. & Newton, R. J. Potential for atmospheric deposition of bacteria to influence bacterioplankton communities. *FEMS Microbiol. Ecol.* **64**, 388–394 (2008).
103. Banat, I. M. *et al.* Thermophilic bacteria in cool temperate soils: are they metabolically active or continually added by global atmospheric transport? *Appl. Microbiol. Biotechnol.* **78**, 841–852 (2008).
104. Perfumo, A. & Marchant, R. Global transport of thermophilic bacteria in atmospheric dust. *Environ. Microbiol. Rep.* **2**, 333–339 (2010).
105. Green, J. L., Womack, A. M. & Bohannon, B. J. M. Biodiversity and biogeography of the atmosphere. *Phil. Trans. R. Soc. B* **365**, 3645–3653 (2010).
106. Fukami, T. In *Community Ecology: Processes, Models and Applications* (eds Verhoef, H. A. & Morin, P. J.) 45–54 (Oxford Univ. Press, 2010).
107. Jessup, C. M. *et al.* Big questions, small worlds: microbial model systems in ecology. *Trends Ecol. Evol.* **19**, 189–197 (2004).
108. Hubert, C. *et al.* A constant flux of diverse thermophilic bacteria into the cold arctic seabed. *Science* **325**, 1541–1544 (2009).
109. Willson, M. F. Dispersal mode, seed shadows, and colonization patterns. *Vegetatio* **108**, 261–280 (1993).
110. Chase, J. M. Drought mediates the importance of stochastic community assembly. *Proc. Natl Acad. Sci. USA* **104**, 17430–17434 (2007).
111. Drake, J. A. Community-assembly mechanics and the structure of an experimental species ensemble. *Am. Nat.* **137**, 1–26 (1991).
112. Fukami, T. & Morin, P. J. Productivity-biodiversity relationships depend on the history of community assembly. *Nature* **424**, 423–426 (2003).
113. Peay, K. G., Belisle, M. & Fukami, T. Phylogenetic relatedness predicts priority effects in nectar yeast communities. *Proc. Biol. Sci.* **279**, 749–758 (2011).
114. Legendre, P. & Legendre, L. *Numerical Ecology (Developments in Environmental Modelling)* (Elsevier, 1998).

#### Acknowledgements

The authors thank A. Martiny, S. Allison, A. Amend, L. Crummett, K. Whiteson, A. Durbin, K. Matulich, S. Kivlin, N. Hynson and J. Talbot for feedback and discussions. This work was supported by a US National Oceanic and Atmospheric Administration–National Estuarine Research Reserve System (NOAA–NERRS) Graduate Research Fellowship to C.A.H. and a US National Science Foundation Grant (OCE-1031783) to J.B.H.M.

#### Competing interests statement

The authors declare no competing financial interests.

#### FURTHER INFORMATION

Jennifer B. H. Martiny's homepage:  
<http://jmartinybio.uci.edu/>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF