Microbial biogeography: putting microorganisms on the map

Jennifer B. Hughes Martiny*, Brendan J.M. Bohannan‡, James H. Brown§, Robert K. Colwell‖, Jed A. Fuhrman¶, Jessica L. Green**, M. Claire Horner-Devine‡‡, Matthew Kane§§, Jennifer Adams Krumins‖‖‖, Cheryl R. Kuske¶¶, Peter J. Morin‖‖‖, Shahid Naeem***, Lise Øvreås‡‡‡, Anna-Louise Reysenbach§§§, Val H. Smith‖‖‖‖‖ and James T. Staley¶¶¶

Abstract | We review the biogeography of microorganisms in light of the biogeography of macroorganisms. A large body of research supports the idea that free-living microbial taxa exhibit biogeographic patterns. Current evidence confirms that, as proposed by the Baas-Becking hypothesis, 'the environment selects' and is, in part, responsible for spatial variation in microbial diversity. However, recent studies also dispute the idea that 'everything is everywhere'. We also consider how the processes that generate and maintain biogeographic patterns in macroorganisms could operate in the microbial world.

Biogeography is the study of the distribution of biodiversity over space and time. It aims to reveal where organisms live, at what abundance, and why. The study of biogeography offers insights into the mechanisms that generate and maintain diversity, such as speciation, extinction, dispersal and species interactions[1]. Since the eighteenth century, biologists have investigated the geographic distribution of plant and animal diversity. More recently, the geographic distributions of microorganisms have been examined. Genetic methodologies have revealed that past culture-based studies missed most microbial diversity[2–4], and have allowed recent studies to sample microbial diversity more deeply and widely than ever before[5,6]. Microbial biogeography stands to benefit tremendously from these advances, although there is still debate as to whether microorganisms exhibit any biogeographic patterns[7–10].

Although traditionally confined to separate academic disciplines, ecologists who study microorganisms and those who study macroorganisms have been interacting more often in recent years. Indeed, this article is a result of a National Center for Ecological Analysis and Synthesis (NCEAS) working group composed of scientists from both specialties. Our goal here is to review what is known regarding the biogeography of microorganisms in light of that of macroorganisms. This inquiry is not driven by the expectation that microorganisms simply follow the patterns of macroorganisms, but rather by the fact that the biogeography of macroorganisms is much better studied. Furthermore, micro- and macroorganisms are often involved in intimate associations that affect each other's geographic distributions[11,12]. Therefore, a logical first hypothesis is that the biogeography of microorganisms is similar to the biogeography of macroorganisms. To the extent that microorganisms conform to the relationships documented for macroorganisms, they will extend the generality of empirical patterns and support mechanistic hypotheses that all living entities share universal attributes. Alternatively, if microorganisms can be shown to represent clear exceptions to the biogeographic patterns of plants and animals, then this will call attention to unique features of microbial life that have influenced the generation and maintenance of its diversity.

As an initial step towards distinguishing between these hypotheses, we suggest a framework for investigating whether microbial assemblages differ in different places and the extent to which this spatial variation is due to contemporary environmental factors and historical contingencies. We then discuss the mechanistic processes that generate and maintain biogeographic patterns in macroorganisms and consider their relevance to microbial biogeography.

There is no universal definition of a 'microorganism'. The term generally denotes members of the domains Bacteria and Archaea, as well as microscopic members of the domain Eukarya (for example, unicellular algae, some fungi and protists). For convenience, we further define a microorganism as having a mass of less than 10^{-5} g and a length of less than 500 μm.

*Department of Ecology and Evolutionary Biology, 80 Waterman Street, BOX G-W, Brown University, Providence, Rhode Island 02912, USA. Correspondence to J.B.H.M. e-mail: jennifer_martiny @brown.edu

Province
A region the biotic composition of which reflects the legacies of historical events.

Habitat type
An environment defined by the suite of its abiotic and biotic characteristics.

We do not consider the question of whether viruses have biogeography, as their biology adds further complications and, in most cases, far less is known about their distribution than that of other microorganisms (for a recent discussion see REF. 13).

A framework for microbial biogeography

A long-standing theme of traditional biogeography is the relative influence of contemporary environmental factors versus the legacies of historical events on present-day distribution patterns. In the early nineteenth century, Augustin P. de Candolle[14] distinguished between the influence of 'habitations' and 'stations' on the distribution of plant diversity. He used the word 'habitation' to signify a biotic province. For instance, the many plant and animal species unique to Australia are attributable to past connections to, and long isolation from, other continents, and clearly distinguish Australia as a distinct province. Augustin P. de Candolle used the word 'station' to mean a habitat type, or a constellation of contemporary abiotic and biotic environmental variables that influenced plant composition. For instance, Australia contains various habitat types that support different biotic assemblages; some habitat types are unique to the province (such as the Mallee scrublands), whereas others are found in many provinces (such as the coastal scrub habitat that has analogues in California, Chile, South Africa and the Mediterranean).

Continents are especially clear examples of macroorganism provinces, but we use the term 'province' more freely than plant and animal provinces have been traditionally defined. A province is any area, the biota of which reflects historical events. Therefore, province size might vary greatly and depend on the particular taxon and resolution of focus. For instance, two lakes a hundred kilometers apart might be separate provinces for a particular strain of bacteria, but all the lakes on a continent might be part of the same province for protist assemblages.

The consideration of habitats and provinces provides a useful framework for addressing four alternative hypotheses. The null hypothesis is that microorganisms are randomly distributed over space. In this case, microorganisms essentially experience only one habitat and one province. A second hypothesis is that the biogeography of microorganisms reflects the influence of contemporary environmental variation (multiple habitats) within a single province. This is the so-called Baas-Becking hypothesis — for microbial taxa, 'everything is everywhere — the environment selects'[15,16]. The claim that 'the environment selects' implies that different contemporary environments maintain distinctive microbial assemblages. The claim 'everything is everywhere' implies that microorganisms have such enormous dispersal capabilities that they rapidly erase the effects of past evolutionary and ecological events. A third alternative is that all spatial variation is due to the lingering effects of historical events (multiple provinces but only one habitat). Historical events that might influence present-day assemblages include dispersal limitation and past environmental conditions, both of which can lead to genetic divergence of microbial assemblages. We consider dispersal limitation to be a historical event, as current composition is influenced by past dispersal limitation, whether relatively recent or ancient. The final hypothesis is that the distributions of microbial taxa, like those of macroorganisms, reflect the influences of both past events and contemporary environmental conditions — in other words, that microbial distributions are shaped by multiple habitats and multiple provinces.

Distinguishing between the four hypotheses addresses two central biogeography questions: first, do microbial assemblages differ in different locations (do microorganisms have biogeography); and second, if microbial assemblages do differ by location, is the spatial variation due to present-day environmental factors, historical contingencies, or both? By definition, differences in microbial assemblages are due to variation in the relative abundances of taxa, including the presence of a particular taxon in one assemblage and its absence in another. As such, we focus on how the relative abundances of microbial taxa vary over space, rather than whether any microbial taxa are truly restricted to particular geographic areas, as it is nearly impossible to conclusively show that a microbial taxon is absent from a given location.

Do microorganisms have biogeography?

It has long been known that many host-associated microorganisms exhibit patterns of genetic, morphological and functional differentiation that are related to the distribution of their hosts[10,12,17–19]. Now, a growing body of evidence shows that free-living microorganisms also vary in abundance, distribution and diversity, over various taxonomic and spatial scales (some examples are given in TABLE 1).

The simplest demonstration of microbial biogeography is that microbial composition across a landscape is non-random, thereby rejecting the first hypothesis above. For example, Cho and Tiedje[20] showed that genetic distance

Author addresses

‡Department of Biological Sciences, Stanford University, California 94305, USA.
§Department of Biology, University of New Mexico, New Mexico 87131, USA.
║Department of Ecology and Evolutionary Biology, University of Connecticut, Connecticut 06269, USA.
¶Department of Biological Sciences, University of Southern California, California 90089, USA.
**School of Natural Sciences, University of California, Merced, California 95344, USA.
‡‡School of Aquatic and Fishery Science, University of Washington, Washington 98195, USA.
§§Division of Molecular and Cellular Biosciences, National Science Foundation, Virginia 22230, USA.
║║Department of Ecology and Evolution, Rutgers University, New Jersey 08901, USA.
¶¶Bioscience Division, Los Alamos National Laboratory, New Mexico 87545, USA.
***Department of Ecology, Evolution and Environmental Biology, Columbia University, New York 10027, USA.
‡‡‡Department of Biology, University of Bergen, Norway.
§§§Department of Biology, Portland State University, Oregon 97207, USA.
║║║Department of Ecology and Evolutionary Biology, University of Kansas, Kansas 66045, USA.
¶¶¶Department of Microbiology, University of Washington, Washington 98195, USA.

Table 1 | **Examples of studies that have found non-random distributions of free-living microbial taxa**

Organisms	Approximate scale (km)	Habitat	Unit	Correlated with	Ref.
Pseudomonads*	20,000	Soil	BOX-PCR isolation	Linear distance	20
3-CBD bacteria‡	20,000	Soil	ARDRA isolation		90
Aerobic, anoxygenic phototrophs‡	20,000	Marine	Dissociation curves	Latitude	91
SAR11 bacteria and archaea‡	13,000	Marine	16S/ITS sequence	Depth	92
Green sulphur bacteria‡	8,000	Lakes	16S sequence	Continental divide	93
N-fixing bacteria‡	700	Desert crusts	Sequence and TRFLP of *nifH* and 16S	Mature versus poorly developed crusts	94
Crenarchaeota*	200	Soil	PCR-SSCP of 16S	At small scales, distance	95
Crenarchaeota*	200	Soil	PCR-SSCP of 16S	Rhizosphere versus bulk soil	96
Bacteria‡	50	Marine	DGGE of 16S	Ocean front	97
Bacteria*	35	Marine	DGGE of 16S	Depth and ocean front	98
Bacteria‡	15	River plume	DGGE of 16S	River–marine transition	99
Bacteria*	5	River plume	DGGE of 16S	Salinity	22
Bacteria, archaea and eukaryotes*	3	Salterns	DGGE, TRFLP, RISA	Salinity	100
Pseudomonas cepacia.*	3	Soil	Isolate allozymes	Vegetation	101
Bacteria and eukaryotes*	1	Soil	RNA hybridization	Cultivation history	102
Gram-negative bacteria*	0.8	Soil	sole carbon source	Latitude	103
Microorganisms*	0.2	Groundwater	RAPD	Oxygen zonation	104
Microorganisms*	0.1	Agricultural soil	AFLP		105
Bacteria and archaea‡	0.02	Lake	DGGE of 16S	Depth	23
Bacteria*	0.01	Drinking water	TRFLP of 16S	Bulkwater versus pipe biofilm	106
Purple non-sulphur bacteria‡	0.01	Fresh marsh	BOX-PCR isolation	Linear distance	21
Bacteria*	0.01	Soil	RFLPs of 16S		28
Microorganisms*	0.002	Salt marsh	RAPD	Marsh elevation	107

The studies are ordered by the geographic scale over which the samples were taken, reported as the approximate furthest distance between sampling points.
*The study found significant non-random distributions. ‡No statistics were performed. If the authors reported that the pattern in microbial composition was correlated with an environmental characteristic, this is reported in the 'correlated with' column, even if this relationship was not statistically tested. 3-CBD, 3-chlorobenzoate-degrading; AFLP, amplified fragment length polymorphism; ARDRA, amplified ribosomal DNA restriction analysis; DGGE, denaturing gradient gel electrophoresis; ITS, intergenic transcribed space; *nifH*, bacterial gene that encodes for nitrogenase; PCR-SSCP, polymerase chain reaction-single strand conformational polymorphism; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphisms; RISA, ribosomal intergenic spacer analysis; TRFLP, terminal RFLP.

between fluorescent pseudomonads was related to geographic distance. Similarly, Oda *et al.*[21] showed genetic differences among purple non-sulphur bacteria along a 10-meter marsh transect. Many of the studies listed in TABLE 1 find correlations between assemblage composition and environmental or geographic characteristics, such as salinity[22], depth[23] and latitude[24].

Taxa–area relationships are further evidence for microbial biogeography. An increase in the number of taxa observed with increasing sample area (often referred to as a species–area relationship) has been detected repeatedly in plants and animals[25]. Recently, investigators have reported similar patterns in microbial communities, in both contiguous[26–28] and island[29,30] habitats. Within contiguous habitats, a positive taxa–area relationship might arise even if microorganisms are randomly distributed over space[31,32]. By contrast, taxa–area relationships that reflect increasing spatial heterogeneity of biotic composition (beta diversity) at increasing spatial scales will exhibit a decrease in biotic similarity with spatial separation — a striking, non-random pattern.

Beta diversity
Taxonomic diversity due to turnover in composition between assemblages.

Distinguishing between environment and history
A limitation of the analyses listed in TABLE 1 is that they exclude only the hypothesis that microbial assemblages are spatially random. This leaves the problem of determining how much of the spatial variation in microbial distributions and assemblages is due to contemporary environmental conditions or historical contingencies. Answering this question requires information on the current abiotic and biotic conditions and the spatial arrangement of the sampled assemblages.

Consider first the case of sampling discrete, pre-defined habitat types that might or might not influence microbial composition. These habitats might be different depths in the ocean water column or rhizospheres of different plant species. As a simple example, consider the case of two distinct geographic locations, each containing three discrete habitat types from which two replicate samples are taken, for a total of 12 sampling sites. The microbial assemblage of each sample is analysed using methods such as clone-library sequence analysis, a community fingerprinting technique or culturing techniques.

Box 1 | Resemblance matrices for biogeographic analyses

Three square resemblance (similarity or distance) matrices are fundamental for biogeographic analyses[35]. These matrices are derived from three data matrices (see figure). The presence/absence or abundance data from all the sites are first summarized in an incidence or abundance matrix, in which letters are the taxa and numbers are the sites in the figure. The taxa are defined by any appropriate operational taxonomic unit (OTU), such as a sequence-similarity cut-off or fingerprint-band length. To calculate the biotic-similarity matrix, the composition is compared between each pair of sites and a similarity index is calculated. The similarity index might be based on presence and absence of each taxon, such as the classic Jaccard index, or also incorporate abundance, such as the Morisita–Horn index[84] or Chao's abundance-based Jaccard estimator[85]. The diagonal entries of the similarity matrix are '1's, and the values above and below the diagonal are mirror images.

The environmental matrix reports the values of each environmental parameter (Roman numerals) recorded at each site and is transformed into an environmental-similarity matrix. One possible similarity index to use is 1 minus a standardized Euclidean distance, in which raw values of environmental variables are first transformed to their standard normal deviate equivalents ([x – mean] divided by the standard deviation) to accommodate the different units of the different variables[27]. In the absence of prior knowledge of which variables influence the microbial community of interest, a large number of factors are often measured. In this case, preliminary analyses are useful to determine the variables that relate to community composition; adding in many unrelated variables can swamp out the signature of any significant variables.

The third matrix is a geographic-distance matrix and is usually the actual geographic distances between each pair of sites, which can be calculated from latitude and longitude values (X and Y). The diagonal values are zero. In some cases, one might want to weight the cell values of a geographic-distance matrix to account for potential barriers to dispersal[86]. For instance, one could account for ocean currents or land masses when investigating marine communities.
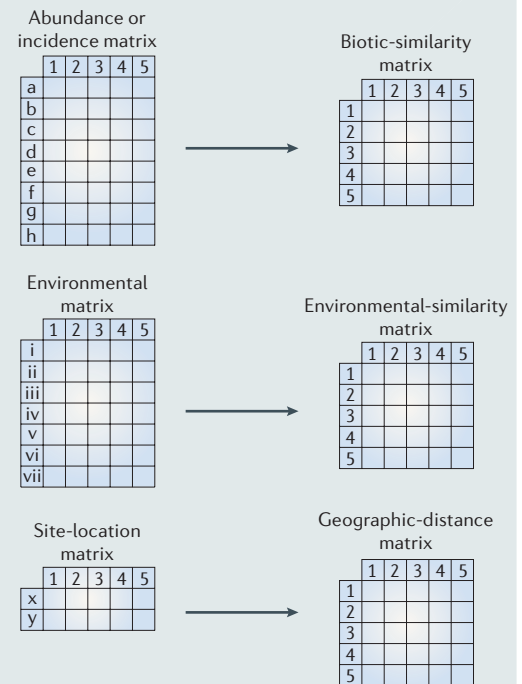
The similarities between each sampling assemblage can be summarized in a biotic-similarity matrix (BOX 1). To picture these data, this matrix can be collapsed with a clustering algorithm. The results of this analysis can then be displayed as a dendrogram[33] or along dimensionless axes with multidimensional scaling[34]. To test the four alternative hypotheses, one can then overlay the information on habitat types and geographic location on the assemblage clustering. FIGURE 1a illustrates an example. The green and white circles represent the two geographic locations. The letters represent different pre-defined habitat types (A, B and C). If the samples are arranged randomly, there is no effect of either current ecology or past history at the taxonomic and spatial resolution sampled, indicating that all the samples were taken within one microbial habitat and one province. Alternatively, biotic composition might cluster into multiple microbial habitats, geographic locations indicating multiple provinces, or both. Various statistical methods can test for significant patterns, such as a two-factor clustering test or canonical analysis[35].

If the samples cluster by habitat, it can be concluded that the assemblages are influenced by the contemporary environment. But what does it mean if geographic separation influences biotic composition? The key to interpreting this result is to be able to determine whether isolation by distance (or a geographic barrier) influences composition even after controlling for present-day environmental factors. Of course, the pre-defined habitats might not capture all possible contemporary environmental variation among sample sites, but with good replication (that is, sampling the same habitat types in many different geographic locations), one can be increasingly certain of a distance effect. Such a distance effect is strong evidence of biogeographic provincialism, in which differences in biotic composition are due to past events rather than present-day attributes of the environment. Although there is no direct effect of distance *per se*, distance is related to the likelihood that past divergence of biotic assemblages, whether due to genetic drift or adaptation to past environments, is maintained by genetic isolation.

Often, researchers record continuous variables to describe the environment and spatial arrangement of their sampling sites rather than using discrete categories. The above analysis can be modified to incorporate continuous measures by deriving two additional matrices: an environmental-similarity matrix (including both biotic and abiotic variables) and a geographic-distance matrix (BOX 1). The correlation, or lack thereof, between the three summary matrices in BOX 1 can then be used to distinguish the four hypotheses (FIG. 1b).

Distance effect
The influence of isolation on biotic composition after controlling for the influence of the contemporary environment.

Genetic drift
Changes in gene frequencies in a population caused solely by chance.

Effects of environment and history

Whereas many studies examine whether microbial assemblages vary over space, we know of only 10 published microbial studies that can be applied to the above framework to assess the roles of historical contingencies
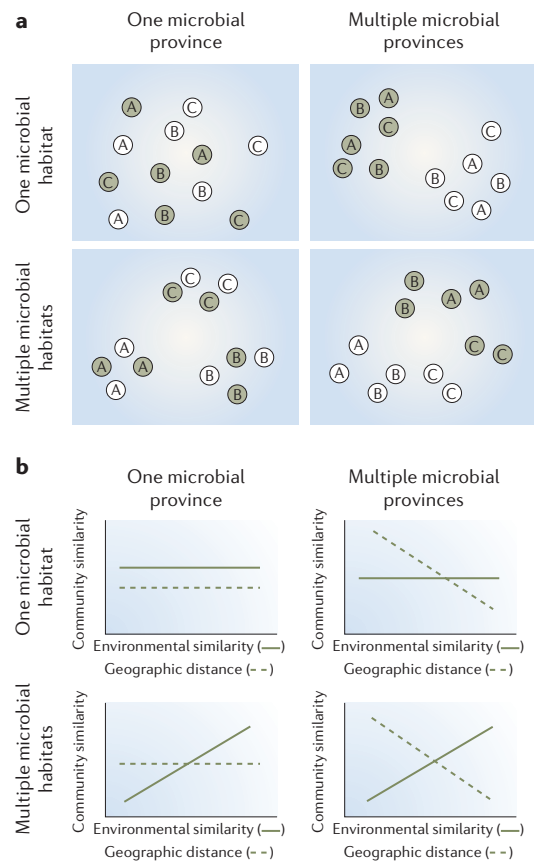

Figure 1 | **Assessing the contributions of environmental and historical effects on microbial biogeography.** Four alternative hypotheses about environmental and historical influences on communities and the general results that would support them, using (**a**) samples (shown in circles) from discrete predefined habitat types (A, B and C) and locations (green versus white) or (**b**) samples from continuous habitat variables and geographic distances. The axes in (**a**) are dimensionless; samples that contain similar assemblages are mapped closer to one another relative to pairs of samples with different assemblages. In **b**, lack of a correlation between environmental similarity or geographic distance and biotic similarity (BOX 1) indicates no biogeographic patterning. Alternatively, to the extent that environmental and historical factors have influenced the assemblages sampled, biotic similarity should be correlated with environmental similarity and geographic distance, respectively. Standard correlation tests are not appropriate to distinguish between these hypotheses because of non-independence; therefore, randomization tests such as a bootstrapped regression analysis[25] or Mantel tests[87,88] are required. Further tests, such as a partial Mantel test, can disentangle the effects of geographic distance versus environment on assemblage composition[89].

and contemporary environmental factors (TABLE 2). The number of studies available is small because few microbial biogeography studies report the geographic distance between their samples or directly test for a distance effect relative to a contemporary environmental effect.

Despite this low number, five studies found significant distance effects, indicating at least some degree of provincialism (TABLE 2). The relative influence of historical versus environmental factors seems to be related to the scale of sampling. In the two intercontinental studies, *Synechococcus*[36] and *Sulfolobus*[37] assemblages in hot springs could be significantly differentiated by distance but did not correlate with the many environmental variables measured. This result indicates that, on the order of tens of thousands of kilometers, the legacy of historical separation can overwhelm any effect of environmental factors. These results should not be taken to mean that the contemporary environment has no effect on biotic composition, but rather that its influence is relatively small compared to that of distance. Indeed, Papke *et al.*[36] noted that, although there was no strong correlation between *Synechococcus* genotype and chemical characteristics of hot springs at an intercontinental scale, some genetic differences among hot springs within a continent seemed to be due to spring chemistry.

By contrast, environmental effects have been repeatedly shown to significantly influence biotic composition at small spatial scales for which distance effects seem to be negligible. The two studies that sampled sites separated by only a few kilometers found significant environmental effects but none of distance (TABLE 2). At intermediate scales (10–3000 km), three of the five studies found a significant distance effect. Environmental conditions also seemed to influence composition at this spatial scale, with one exception[38]. Therefore, it is at this intermediate spatial scale that the influence of both historical contingencies and contemporary ecological factors on microbial biogeography is most likely to be detected.

These general trends were apparent even though the studies included a broad range of taxa (bacteria, archaea and fungi) and the resolution of the taxonomic units varied enormously. The OTU (operational taxonomic unit) definitions varied from ARISA (automated ribosomal intergenic spacer analysis) profiles of all bacteria to multilocus sequence typing of cultured isolates within one genus. In general, patterns present at finer taxonomic resolutions might not be reflected at broader resolutions. For instance, one might not see compositional differences among sites at the level of 16S rDNA, even though there are clear differences in ITS (intergenic transcribed space) sequences[20]. The same taxonomic-scale dependence applies to the distribution of macroorganisms; for instance, many more plant genera are restricted to a particular continent than plant families.

What processes shape microbial biogeography?

The studies reviewed above indicate that microbial assemblages can exhibit both environmental segregation and biogeographic provincialism (TABLE 2), but what processes generate these patterns? The definitive

Table 2 | **Studies of the effects of distance (dist.) and environment (env.) on microbial composition**

Organisms	Approximate scale (km)	Habitat	OTU	Effect of dist.	env.	Ref.
Synechococcus	20,000	Hot springs	16S/ITS sequence	Yes	No	36
Sulfolobus	12,000	Hot springs	MLS of isolates	Yes*	No*	37
Bacteria	3,000	Coral	16S sequence	No	Yes*	108
Bacteria	500	Lakes	ARISA	Yes*	Yes*	109
3-CBD bacteria	500	Soil	ARDRA	No	Yes*	87
Ascomycetes	100	Soil	ARISA	Yes*	Yes*	26
Bacteria	100	Aquatic	ARISA	No	Yes	110
Bacteria	10	Lakes	DGGE of 16S	Yes*	No*	38
Bacteria	0.3	Marsh sediment	16S sequence	No*	Yes*	27
Bacteria	0.1	Soil	TRFLP	No	Yes*	33

The studies are ordered by the geographical scale over which the samples were taken, reported as the approximate furthest distance between sampling points. *The effect was tested for statistical significance. 3-CBD, 3-chlorobenzoate-degrading; ARDRA, amplified ribosomal DNA restriction analysis; ARISA, automated ribosomal intergenic spacer analysis; DGGE, denaturing gradient gel electrophoresis; ITS, intergenic transcribed space; MLS, multilocus sequencing; OTU, operational taxonomic unit used in the study; TRFLP, terminal RFLP.

difference between all micro- and macroorganisms is their size. As a first hypothesis, we suggest that the same processes that influence macroorganism biogeography also apply to microbial life but that their rates scale with body size or, for single-celled organisms, cell size. Many attributes, from metabolic rate to maximum lifespan, vary predictably with an organism's size. Often, this variation yields linear relationships on a logarithmic plot and therefore can be described by a so-called allometric equation. This idea of allometry[39] serves as a useful structure for the discussion of biogeographic processes below.

For plant and animal biogeography, the most relevant rates are those processes by which a taxon expands or contracts its area of distribution[1]. We discuss three such processes: colonization, speciation (generally, diversification) and extinction.

Colonization. One of the main arguments behind the 'everything is everywhere' hypothesis of microbial biogeography is that the dispersal and subsequent colonization of microorganisms into new locations is so great that it prevents spatial differentiation. High dispersal rates decrease assemblage differentiation by increasing gene flow, whether by means of sexual reproduction or horizontal gene transfer, overwhelming any tendency towards genetic differentiation due to mutation, selection or genetic drift[1]; and by mixing of individuals, overwhelming spatial differences in taxon abundance due to ecological interactions or to ecological drift[40,41].

From the perspective of allometry, the question is: does an organism's size influence its dispersal ability? Of particular interest are long-distance dispersal events that transport a propagule across barriers of inhospitable habitats. FIGURE 2 shows the hypothesized relationship between body mass and dispersal capability, which is the maximal distance traveled by an individual in its lifetime (or between cell divisions). This relationship depends on whether an organism disperses primarily by active propulsion, such as propelling itself micrometers with its flagella, or by passive transport, such as being carried thousands of kilometers by ocean currents or migrating birds.

Several points are immediately apparent from FIG. 2. First, dispersal capacity is one of the least likely attributes to be constrained by the size of an organism and is not well characterized by an allometric equation. We can, however, hypothesize about the potential constraints of this relationship. Second, there are severe constraints on microbial dispersal by active propulsion (FIG. 2a). Large organisms range from having little or no active dispersal (trees, giant clams and corals) to dispersing over
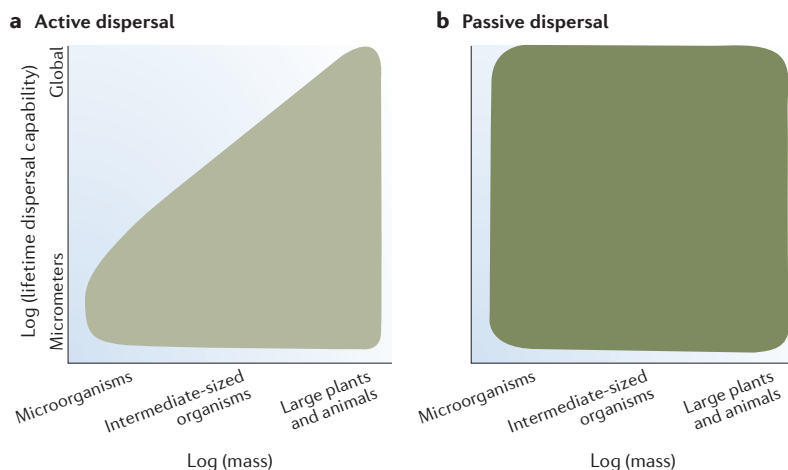
a Active dispersal

b Passive dispersal

Figure 2 | **Hypothetical relationship between body mass (at an organism's largest life stage) and lifetime dispersal capability.** The relationship varies depending on whether the organism disperses actively or passively (by its own propulsion). The range of active (**a**) dispersal is a subset of the range of passive (**b**) dispersal. It is convenient to think of the log(mass) axis as representing three qualitative groups: first, microorganisms, which span about 8 orders of magnitude from bacteria to eukaryotic algae and protozoa (10^{-13}–10^{-5} g); second, large plants and animals, which span about 8 orders of magnitude from herbs and small vertebrates to whales and trees (10^1–10^9 g); and third, intermediate-sized organisms, which span the intervening 6 or so orders of magnitude and include the small metazoans, such as nematodes, annelids and arthropods.

thousands of kilometers (whales, birds and butterflies). By contrast, microorganisms have little capacity to cross significant geographic barriers under their own propulsion. Of course, over many generations, a bacterial taxon could eventually spread great distances by active propulsion. By that time, however, genetic divergence from the source population would likely occur, thereby generating biogeographic structure rather than eliminating it. Third, there seem to be no size constraints on passive dispersal (FIG. 2b). Some large organisms, such as elephants and rhinoceroses, have negligible passive dispersal, whereas others, such as tree ferns, trees and giant clams, disperse thousands of kilometers by air or water as spores, seeds and larvae. We extrapolate that passive dispersal by microorganisms is equally broad — whereas some microbial taxa might disperse globally, others will only disperse over very short distances, creating non-random distributions of microbial assemblages.

What factors might limit the passive dispersal of a microorganism (and fill in the bottom-left-hand corner of FIG. 2b)? Certainly, habitat will have a role. Cells in subsurface soils and sediment will not disperse as far as those in water and surface soils. Furthermore, the propagule must survive the conditions encountered during dispersal to a new suitable location. The 'everything is everywhere' dictum implicitly assumes that all microorganisms are highly tolerant to stress; however, not all microorganisms produce spores and cysts, and those that do vary greatly in their hardiness. Last, to grow and establish a population, the propagule must be able to outcompete local populations that might be better adapted to the specific conditions.

A taxon's colonization rate depends on its population density as well as its dispersal ability. This relationship is best illustrated by a dispersal frequency distribution (FIG. 3). For passively dispersed macroorganisms,

Figure 4 | **Hypothesized constraints on a taxon's population density in a given body-size class.** The thick green line on the diagonal is a known physiological constraint. The gradient in shading from the diagonal to the bottom-left corner represents the idea that fewer taxa are thought to fall in the bottom left of the figure; however, we hypothesize that some taxa do fall in this region. The inset plots log (body mass, g) of North American birds versus log (population density, individuals per route). The data set falls within a well-defined quadrilateral with a constant minimum density and a maximum density for birds of an intermediate size. Individual data points are not shown. The outline of these data is also sketched on the constraint figure. The approximate range of marine phytoplankton data from Li[48] is also sketched (assuming that cell volume is proportional to body mass). The X-axis categories are defined in FIG. 2. Inset adapted with permission from REF. 43 © (1987) University of Chicago.
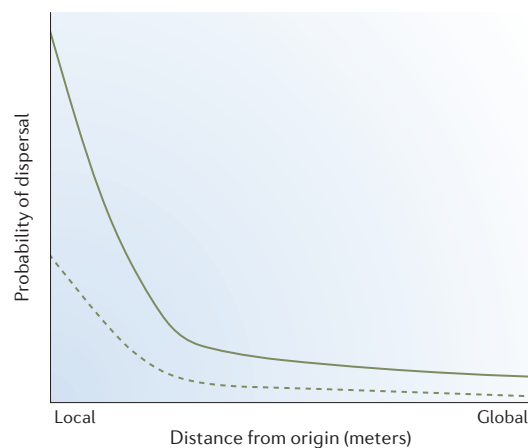

Figure 3 | **Hypothetical dispersal distribution of a typical passively dispersed macroorganism.** Population density influences the probability that an individual from that population will disperse over very long distances (solid line). For taxa with relatively low densities (dashed line), dispersal might be effectively restricted, even though long-distance movement is theoretically possible. Based on REF. 41.
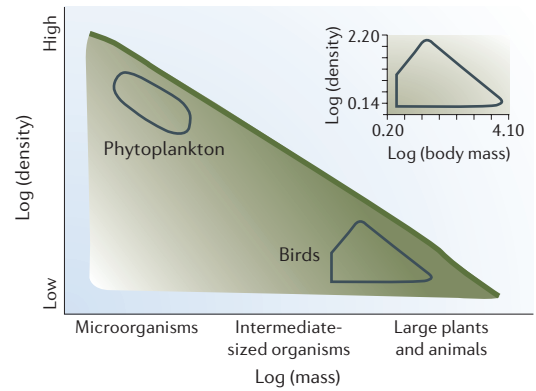
most propagules move only very short distances, but a small proportion can disperse over vast distances (for example, in seed dispersal)[42]. As similar processes contribute to their dispersal, we hypothesize that the shape of the frequency distribution of microbial dispersal distances is similar to that observed with passively dispersed macroorganisms. In general, large numbers of potential propagules increase the chance that at least one will travel a long distance and establish a new persistent population, whereas low densities effectively shorten the tail of the dispersal distribution (FIG. 3). Given that microorganisms have finite population sizes, the low-probability, long-distance dispersal events that are expected to occur eventually by chance might occur rarely or not at all.

How do the population densities of microbial taxa compare to those of macroorganisms? FIGURE 4 illustrates the relationship observed in macroorganisms and extends these observations to microorganisms. There are three important features to note here. First, on average, there seems to be a negative relationship between a taxon's size and its population density. In many studies[43–46], smaller-sized organisms have, on average, larger population densities than bigger organisms (although the maximum population density is often not the smallest size category)[47]. If the same relationship holds at the scale of microbial sizes, then, on average, microorganisms will have larger population densities than macroorganisms. Indeed, Li[48] found a negative relationship between the

cell volume of marine phytoplankton and their population densities (FIG. 4). Second, there is no reason to think that small organisms that have low population densities do not exist (in other words, there are taxa in the bottom-left-hand corner of FIG. 4). Indeed, this must be true if, for no other reason, the population density of a microbial taxon must be low following a diversification event or just prior to extinction. Last, there are theoretical and empirical reasons to expect that the upper limit to population density for a given size class increases as size decreases (the thick line in FIG. 4). Smaller organisms need fewer resources per individual and can therefore have higher population densities in an area.

In conclusion, the combination of dispersal ability and population density of a taxon determines its rate of colonizing new and distant habitats. Whereas some microbial taxa could possess the combination of traits that allows them to colonize at a global scale (spore-forming *Bacillus*[49], for example), others might have short dispersal distances and restricted geographic distributions (hot-spring *Sulfolobus*[37], for example).

Diversification and extinction. Diversification owing to mutation, genetic drift and differential selective pressures will generate biogeographic patterns, unless it is counterbalanced by the forces of dispersal and homogenizing selection. Evidence is mixed as to whether speciation rates are related to body size in macroorganisms[50,51]. If laboratory studies are any indication[52,53], the potential for rapid diversification seems greater in microorganisms than macroorganisms. Microbial species typically have higher densities (FIG. 4) and shorter generation times than macroorganisms, allowing them to undergo rapid genetic divergence.

Extinction influences biogeographic patterns, usually by eliminating endemic forms. Species with small geographic-range sizes have a higher probability of extinction[54,55]; large geographic ranges provide insurance against extinction owing to local disturbances. In macroorganisms, there is a weak but significant positive correlation between body size and range size (FIG. 5), and the relationship is better characterized as a constraint space. The upper limit on range size extends to the entire globe and is independent of body mass, as examples of global taxa are found among bacteria[49] as well as plants and animals[1,56]. By contrast, the lower limit of range size probably depends on organism size. There are few large organisms that occupy small ranges (FIG. 5), probably because the lower population density of larger organisms (FIG. 4) results in an increased likelihood of extinction[43,57]. Indeed, many large-bodied mammals and birds that occupy limited ranges are recently extinct or are currently endangered[58].

How might range sizes vary within the constraints of FIG. 5, and therefore influence microbial extinction rates? The fact that there are some 'cosmopolitan' microorganisms (for examples, see REFS 59-61) should not be taken to imply that most taxa are so widely distributed. The modal range size of macroorganisms within a taxonomic group tends to be intermediate — that is, most species are neither extremely narrowly nor globally distributed[62]. If the pattern in macroorganisms extends to microorganisms, we

expect that range sizes within microbial-cell size classes vary greatly. Variation in range sizes indicates variation in extinction rates; those microorganisms that have relatively restricted distributions should have higher extinction rates than those that are more global in range.

The balance of biogeographic processes. Ultimately, it is a balance between origination and extinction processes that determines global taxon diversity and shapes biogeography. There is evidence that the balance of global species diversity and, therefore, the underlying processes scale with organism size (FIG. 6; for examples, see REFS 63,64). The exact relationship is still unclear, however, and the number of species might decrease with decreasing size at the lower limit of the size distribution of macroorganisms. This decrease might be an artefact owing to undersampling and taxonomic lumping of small organisms or, alternatively, it could be the true relationship.

These two alternatives predict large differences in total microbial diversity. If we assume that the apparent decrease in richness of the smaller macroorganisms is an artefact and that the pattern for larger macroorganisms reflects the true relationship, then microbial diversity within a size class is predicted to increase allometrically with decreasing size. As such, microbial diversity would be much higher than the diversity of macroorganisms (as suggested by Dykhuizen[65]). By contrast, if richness peaks at some intermediate body size (as suggested in FIG. 6), then total microbial diversity might be lower than total macroorganism diversity (as suggested by Finlay[9]). Of course, a complication in comparing the biogeography of micro- and macroorganisms is the problem of comparing equivalent taxonomic units. Not only would sufficient sampling be needed to assess the global diversity of microorganisms within different size classes, but their diversity would need to be measured using taxon definitions comparable to those of plants and animals.
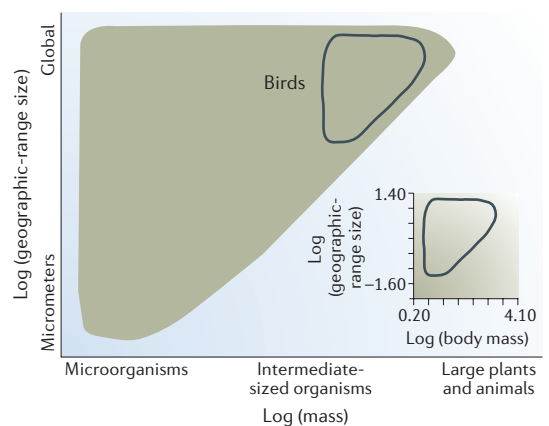

Figure 5 | **Hypothesized constraints on an organism's geographic-range size for a given body mass.** The inset graph is log (body size, g) versus log (geographic-range size, 10^6 km^2) for terrestrial bird species of North America. Individual data points are not shown. The combined data set forms an approximate triangle. The outline of these data is also sketched on the constraint figure. The X-axis categories are defined in FIG. 2. Inset adapted with permission from REF. 43 © (1987) University of Chicago.

Geographic range
The area encompassing the extent of a taxon's distribution.

Determining the nature of this relationship would provide insight into the nature of the processes underlying biogeographic patterns. If the relationship between global diversity and body size proves to be hump-shaped, this indicates that diversification and extinction do not both scale allometrically with body size, and that a body-size threshold might exist across which the balance between diversification and extinction is determined by fundamentally different factors.

Discussion

A large body of research supports the idea that free-living microorganisms exhibit biogeographic patterns. Current evidence confirms that, as Baas-Becking proposed, the environment selects and is, in part, responsible for spatial variation in microbial diversity. However, recent studies dispute the idea that 'everything is everywhere'. Instead, the legacies of historical events have left lasting signatures on the distributions of microbial assemblages, even at distances as small as 500 km.

These results indicate that there are some aspects of biogeography that might be common to all of life. However, there are other aspects of biogeography that might be unique to microorganisms. For example, we conclude that the rates of the processes underlying biogeography probably vary more widely for microorganisms of a given size than for macroorganisms of a given size. Except for the case of active dispersal, we hypothesize that body size does not constrain a microorganism's dispersal rate, population density and range size, whereas it does somewhat constrain those of a larger organisms (FIGS 2,4,5). Therefore, a question for future research is: what are the traits that lead to the wide variety of colonization, diversification and extinction rates in microorganisms (and which at the same time

are relatively more constrained for macroorganisms)? Our discussion was limited by treating biogeographic processes separately; further work is needed to assess the relative importance of these processes for different types of microorganisms.

An unavoidable problem in comparing the patterns of micro- and macroorganisms is how to compare equivalent taxonomic units. In theory, the same taxonomic resolution should be used, whether that unit is a species or a sequence-similarity cut-off. In practice, most macroorganism taxa, species or otherwise are defined by morphological characteristics, which are not tightly correlated with genetic differentiation. For instance, the degree of genetic variation of fish, bird and mammal species within genera spans two orders of magnitude among these genera[66].

Furthermore, the level of taxonomic resolution in microbial-diversity studies is generally much coarser than that adopted with macroorganisms. For example, human and chimpanzee genomes exhibit 98.6% DNA–DNA hybridization between them, yet bacterial systematists have traditionally defined a species as strains with genomes that exhibit over 70% DNA–DNA hybridization[8]. Similarly, microbial eukaryotes are often differentiated by morphological characteristics, but their small size means that few characteristics can be distinguished. Therefore, a single Latin binomial for a microbial eukaryote can often refer to a complex of cryptic species[67]. Consequently, declaring that these 'species' are cosmopolitan, for example[7,9], might be approximately equivalent to saying that a genus or family of birds is cosmopolitan.

Just as the long-standing debate about species definitions remains unresolved for microorganisms[68] and macroorganisms[69], so will the broader question of whether, and how, to compare them. We are optimistic, however. Despite the difficulty of defining microbial taxonomic units, biogeographic patterns seem robust enough to be detectable across various taxa (TABLE 1). It is also possible to avoid comparing taxonomic units by asking instead whether there is a level of taxonomic resolution at which microbial biogeographic patterns approach those of macroorganisms[70].

The four hypotheses illustrated in FIG. 1 provide a useful framework to further explore microbial biogeography, beyond merely documenting the existence of patterns. Many microbial data sets have already been collected that could be interpreted within this framework. Many, however, cannot, primarily because such studies have failed to take samples separated by a range of known distances, to sample across various habitats, to gather contemporaneous environmental data or all of the above. In general, designing and organizing data sets to correlate microbial genomic and metagenomic data with ecological and environmental information can provide biological insights[71]. Specifically, we recommend that new microbial biogeography studies should systematically sample and record data from various distances, habitats and environmental conditions, to better distinguish between contemporary and historical factors. If they do not, the field of microbial biogeography will probably become mired in phenomenological description, instead of tackling the mechanisms that generate the patterns.
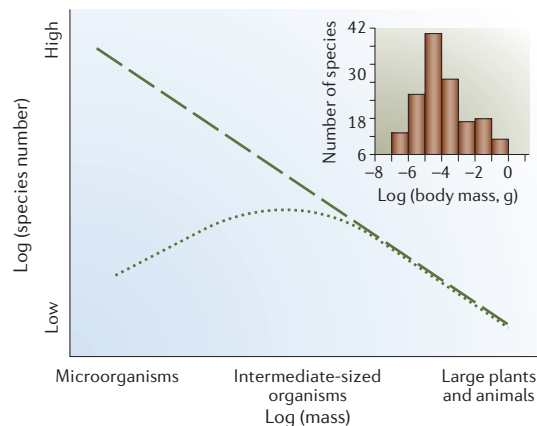

Figure 6 | **Hypothesized relationships between number of species and body mass.** For larger macroorganisms (approximately larger than insects), it is clear that the number of species increases as body mass decreases. For smaller macroorganisms and microorganisms, the number of species might continue to increase (dashed line) or begin to decrease (dotted line) as body mass decreases. The inset plots the number of invertebrate species by log(mass) on Marion Island[64]. The X-axis categories are defined in FIG. 2. Inset reproduced with permission from REF. 64 © (2001) National Academy of Sciences, USA.

The motivation for understanding microbial biogeography extends beyond drawing and interpreting a map of microbial diversity. As with macroorganisms (for example, see REFS 72–75), a growing body of evidence indicates that microbial composition also affects ecosystem processes, including CO_2 respiration and decomposition[76,77], autotrophic and heterotrophic production[78,79] and nitrogen cycling[80,81]. Therefore, even under similar environmental conditions, microbial communities from different provinces might function differently. A better understanding of microbial biogeography is essential to predict such effects. It is also crucial in the search for novel pharmaceuticals and other compounds of industrial importance[82].

If our initial biogeographic hypotheses based on extrapolating from macroorganisms to microorganisms are eventually rejected, this outcome would support the view that "biodiversity at the microbial level is fundamentally different from that of macroscopic animals and plants"[83]. Discovering exactly which attributes and processes contribute to these fundamental differences would greatly further our understanding of all living things. If, however, biogeographic and allometric patterns and processes are found to be fundamentally similar in all organisms, this will provide yet another example of the unity of life — the extent to which all living things not only use similar molecules for structure and function, but also follow similar ecological, evolutionary and biogeographic principles.

1. Brown, J. H. & Lomolino, M. V. *Biogeography* (Sinauer, Sunderland, 1998).
   A definitive textbook on the biogeography of macroorganisms; has just been updated in a 2005 edition.
2. Ward, D. M., Weller, R. & Bateson, M. M. 16S ribosomal-RNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**, 63–65 (1990).
3. Øvreås, L. Population and community level approaches for analysing microbial diversity in natural environments. *Ecol. Lett.* **3**, 236–251 (2000).
4. Floyd, M. M., Tang, J., Kane, M. & Emerson, D. Captured diversity in a culture collection: case study of the geographic and habitat distributions of environmental isolates held at the American type culture collection. *Appl. Environ. Microbiol.* **71**, 2813–2823 (2005).
5. Schloss, P. D. & Handelsman, J. Status of the microbial census. *Microbiol. Mol. Biol. Rev.* **68**, 686–691 (2004).
6. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
7. Fenchel, T., Esteban, G. F. & Finlay, B. J. Local versus global diversity of microorganisms: cryptic diversity of ciliated protozoa. *Oikos* **80**, 220–225 (1997).
8. Staley, J. T. Biodiversity: are microbial species threatened? *Curr. Opin. Biotechnol.* **8**, 340–345 (1997).
9. Finlay, B. J. Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063 (2002).
   A summary of the arguments for why microbial eukaryotes might not be restricted by geographic barriers.
10. Hedlund, B. P. & Staley, J. T. Microbial endemism and biogeography. In *Microbial Diversity and Bioprospecting* (ed. Bull, A. T.) (ASM, Washington DC, 2003).
11. Anagnostakis, S. Chestnut blight: the classical problem of an introduced pathogen. *Mycologia* 79, 23–27 (1987).
12. Falush, D. *et al.* Traces of human migrations in *Helicobacter pylori* populations. *Science* 299, 1582–1585 (2003).
13. Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* 13, 278–284 (2005).
14. de Candolle, A. P. *Essai Elementaire de Geographie Botanique* (F. G. Levrault, Paris, 1820).
15. Beijerinck, M. W. De infusies en de ontdekking der backteriën. In *Jaarboek van de Koninklijke Akademie van Wetenschappen* (Müller, Amsterdam, 1913).
16. Baas-Becking, L. G. M. *Geobiologie of Inleiding Tot de Milieukunde* (Van Stockkum & Zoon, The Hague, 1934).
17. Zhang, N. & Blackwell, M. Population structure of dogwood anthracnose fungus. *Phytopathology* 92, 1276–1283 (2002).
18. Bala, A., Murphy, P. & Giller, K. E. Distribution and diversity of rhizobia nodulating agroforestry legumes in soils from three continents in the tropics. *Mol. Ecol.* 12, 917–929 (2003).
19. Papke, R. T. & Ward, D M. The importance of physical isolation to microbial diversification. *FEMS Microbiol. Ecol.* 48, 293–303 (2004).
   A recent review discussing how physical isolation might affect prokaryote evolution.
20. Cho, J. C. & Tiedje, J. M. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* 66, 5448–5456 (2000).
    One of the first examples of relating the genetic similarity of a free-living bacterial assemblage with geographic distance, using fluorescent *Pseudomonas* isolates.
21. Oda, Y., Star, B., Huisman, L. A., Gottschal, J. C. & Forney, L. J. Biogeography of the purple nonsulfur bacterium *Rhodopseudomonas palustris*. *Appl. Environ. Microbiol.* **69**, 5186–5191 (2003).
22. Crump, B. C., Hopkinson, C. S., Sogin, M. L. & Hobbie, J. E. Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. *Appl. Environ. Microbiol.* **70**, 1494–1505 (2004).
23. Øvreås, L., Forney, L., Daae, R. L. & Torsvik, V. Distribution of bacterioplankton in meromictic Lake Sælenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl. Environ. Microbiol.* **63**, 3367–3373 (1997).
24. Schwalbach, M. S., Hewson, I. & Fuhrman, J. A. Viral effects on bacterial community composition in marine plankton microcosms. *Aquat. Microb. Ecol.* **34**, 117–127 (2004).
25. Rosenzweig, M. L. *Species Diversity in Space and Time*. (Cambridge University Press, Cambridge, 1995).
26. Green, J. L. *et al.* Spatial scaling of microbial eukaryote diversity. *Nature* **432**, 747–750 (2004).
    Along with reference 27, this study uses distance–decay curves to demonstrate non-random bacterial-assemblage distributions; the studies then test whether the non-random distributions are due to isolation by distance and/or local environmental conditions.
27. Horner-Devine, M. C., Lage, M., Hughes, J. B. & Bohannan, B. J. M. A taxa–area relationship for bacteria. *Nature* **432**, 750–753 (2004).
28. Noguez, A. M. *et al.* Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Glob. Ecol. Biogeogr.* **14**, 241–248 (2005).
29. Bell, T. *et al.* Larger islands house more bacterial taxa. *Science* **308**, 1884 (2005).
30. Smith, V. H. *et al.* Phytoplankton species richness scales consistently from laboratory microcosms to the world's oceans. *Proc. Natl Acad. Sci. USA* **102**, 4393–4396 (2005).
31. Cam, E. *et al.* Disentangling sampling and ecological explanations underlying species–area relationships. *Ecology* **84**, 1118–1130 (2002).
32. Colwell, R. K., Mao, C. X. & Chang, J. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* **85**, 2717–2727 (2004).
33. Kuske, C. R. *et al.* Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Appl. Environ. Microbiol.* **68**, 1854–1863 (2002).
34. Yannarell, A. C. & Triplett, E. W. Within- and between-lake variability in the composition of bacterioplankton communities: investigations using multiple spatial scales. *Appl. Environ. Microbiol.* **70**, 214–223 (2004).
35. Legendre, P. & Legendre, L. *Numerical Ecology* 2nd edn (Elsevier, Amsterdam, 1998).
36. Papke, R. T., Ramsing, N. B., Bateson, M. M. & Ward, D. M. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**, 650–659 (2003).
37. Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic Archaea. *Science* **301**, 976–978 (2003).
    Already a classic microbial biogeography study, it considers both the effects of spatial isolation and contemporary environmental parameters on hotspring *Sulfolobus* assemblages within and across continents.
38. Reche, I., Pulido-Villena, E., Morales-Baquero, R. & Casamayor, E. O. Does ecosystem size determine aquatic bacterial richness? *Ecology* **86**, 1715–1722 (2005).
39. Brown, J. H., West, G. B. & Enquist, B. J. Scaling in biology: patterns and processes, causes and consequences. In *Scaling in Biology* (eds Brown, J. H. & West, G. B.) 1–24 (Oxford University Press, Oxford, 2000).
    A comprehensive introduction to allometric patterns in biology.
40. Bell, G. The distribution of abundance in neutral communities. *Am. Nat.* **155**, 606–617 (2000).
41. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, 2001).
42. Clark, J. S., Silman, M., Kern, R., Macklin, E. & HilleRisLambers, J. Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology* **80**, 1475–1494 (1999).
43. Brown, J. H. & Maurer, B. A. Evolution of species assemblages: effects of energetic constraints and species dynamics on the diversification of the North American avifauna. *Am. Nat.* **130**, 1–17 (1987).
44. Morse, D. R., Stork, N. E. & Lawton, J. H. Species number, species abundance and body length relationships of arboreal beetles in Bornean lowland rain forest trees. *Ecol. Entomol.* **13**, 25–37 (1988).
45. Lawton, J. H. Species richness and population dynamics of animal assemblages. Patterns in body size: abundance space. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **330**, 283–291 (1990).
46. Siemann, E., Tilman, D. & Haarstad, J. Insect species diversity, abundance and body size relationships. *Nature* **380**, 704–706 (1996).
47. Brown, J. H. *Macroecology* (Chicago University Press, Chicago, 1995).
48. Li, W. K. W. Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. *Nature* **419**, 154–157 (2002).
49. Roberts, M. S. & Cohan, F. M. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* **49**, 1081–1094 (1995).
50. Stanley, S. M. *Macroevolution: Pattern and Process* (W. H. Freeman, San Francisco, 1979).
51. Paradis, E. Statistical analysis of diversification with species traits. *Evolution* **59**, 1–12 (2005).
52. Lenski, R. E. & Travisano, M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial-populations. *Proc. Natl Acad. Sci. USA* **91**, 6808–6814 (1994).

53. Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998).

54. Manne, L. L., Brooks, T. M. & Pimm, S. L. Relative risk of extinction of passerine birds on continents and islands. *Nature* **399**, 258–261 (1999).

55. Purvis, A., Gittleman, J. L., Cowlishaw, G. & Mace, G. M. Predicting extinction risk in declining species. *Proc. R. Soc. Lond. B Biol. Sci.* **267**, 1947–1952 (2000).

56. Lawton, J. H. Range, Population abundance and conservation. *Trends Ecol. Evol.* **8**, 409–413 (1993).

57. Silva, M. & Downing, J. A. Allometric scaling of minimal mammal densities. *Conserv. Biol.* **8**, 732–743 (1994).

58. Meretsky, V. J., Snyder, N. F. R., Beissinger, S. R., Clendenen, D. A. & Wiley, J. W. Demography of the California condor: implications for reestablishment. *Conserv. Biol.* **14**, 957–967 (2000).

59. Finlay, B. J. & Clarke, K. J. Apparent global ubiquity of species in the protist genus *Paraphysomonas*. *Protist* **150**, 419–430 (1999).

60. Massana, R., DeLong, E. F.& Pedrós-Alió, C. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl. Environ. Microbiol.* **66**, 1777–1787 (2000).

61. Lindström, E. S. & Leskinen, E. Do neighboring lakes share common taxa of bacterioplankton? Comparison of 16S rDNA fingerprints and sequences from three geographic regions. *Microb. Ecol.* **44**, 1–9 (2002).

62. Brown, J. H., Stevens, G. C. & Kaufman, D. M. The geographic range: size, shape, boundaries, and internal structure. *Annu. Rev. Ecol. Syst.* **27**, 597–623 (1996).

63. May, R. M. How many species are there on Earth? *Science* **241**, 1441–1449 (1988).

64. Gaston, K. J., Chown, S. L. & Mercer, R. D. The animal species-body size distribution of Marion Island. *Proc. Natl Acad. Sci. USA* **98**, 14493–14496 (2001).

65. Dykhuizen, D. E. Santa Rosalia revisited: why are there so many species of bacteria? *Antonie Van Leeuwenhoek* **73**, 25–33 (1998).

66. Avise, J. C. & Aquadro, C. F. A comparative summary of genetic distances in the vertebrates. *Evol. Biol.* **15**, 151–185 (1982).

67. de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W. & Pawlowski, J. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc. Natl Acad. Sci. USA* **96**, 2864–2868 (1999).

68. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487 (2002).

69. Coyne, J. A., Orr, H. A. & Futuyma, D. J. Do we need a new species concept? *Syst. Zool.* **37**, 190–200 (1988).

70. Green, J. L. & Bohannan, B. J. M. Spatial scaling of microbial biodiversity in *Scaling Biodiversity* (eds Storch, D. & Marquet, P. A. & Brown, J. H.) (Cambridge University Press, Cambridge, 2006).

71. Martiny, J. B. H. & Field, D. Ecological perspectives on the sequenced genome collection. *Ecol. Lett.* (in the press).

72. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).

73. Hooper, D. U. & Vitousek, P. M. The effects of plant composition and diversity on ecosystem processes. *Science* **277**, 1302–1305 (1997).

74. Dukes, J. S. Biodiversity and invasibility in grassland microcosms. *Oecologia* **126**, 563–568 (2001).

75. Treseder, K. K. & Vitousek, P. M. Potential ecosystem-level effects of genetic variation among populations of *Metrosideros polymorpha* from a soil fertility gradient in Hawaii. *Oecologia* **126**, 266–275 (2001).

76. McGrady-Steed, J., Harris, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165 (1997).

77. Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L. & Lilley, A. K. The contribution of species richness and composition to bacterial services. *Nature* **436**, 1157–1160 (2005).

78. Naeem, S. & Li, S. B. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509 (1997).

79. van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).

80. Cavigelli, M. A. & Robertson, G. P. The functional significance of denitrifier community composition in a terrestrial ecosystem. *Ecology* **81**, 1402–1414 (2000).

81. Horz, H.-P., Barbrook, A., Field, C. B. & Bohannan, B. J. M. Ammonia-oxidizing bacteria respond to multifactorial global change. *Proc. Natl Acad. Sci. USA* **101**, 15136–15141 (2004).

82. Bull, A. T., ed. *Microbial Diversity and Bioprospecting*. (ASM Press, Washington DC, 2003).

83. Finlay, B. J., Esteban, G. F., Olmo, J. L. & Tyler, P. A. Global distribution of free-living microbial species. *Ecography* **22**, 138–144 (1999).

84. Magurran, A. E. *Ecological Diversity and Its Measurement* (Princeton University Press, Princeton, 1988).

85. Chao, A., Chazdon, R. L., Colwell, R. K. & Shen, T. J. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol. Lett.* **8**, 148–159 (2005).

86. Ricketts, T. H. The matrix matters: effective isolation in fragmented landscapes. *Am. Nat.* **158**, 87–99 (2001).

87. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).

88. Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055 (1992).

89. Smouse, P. E., Long, J. C. & Sokal, R. R. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* **35**, 627–632 (1986).

90. Fulthorpe, R. R., Rhodes, A. N. & Tiedje, J. M. High levels of endemicity of 3-chlorobenzoate-degrading soil bacteria. *Appl. Environ. Microbiol.* **64**, 1620–1627 (1998).

91. Schwalbach, M. S. & Fuhrman, J. A. Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and quantitative PCR. *Limnol. Oceanogr.* **50**, 620–628 (2005).

92. Garcia-Martinez, J. & Rodriguez-Valera, F. Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I. *Mol. Ecol.* **9**, 935–948 (2000).

93. Glaeser, J. & Overmann, J. Biogeography, evolution, and diversity of epibionts in phototrophic consortia. *Appl. Environ. Microbiol.* **70**, 4821–4830 (2004).

94. Yeager, C. M. *et al.* Diazotrophic community structure and function in two successional stages of biological soil crusts from the Colorado plateau and Chihuahuan desert. *Appl. Environ. Microbiol.* **70**, 973–983 (2004).

95. Sliwinski, M. K. & Goodman, R. M. Spatial heterogeneity of Crenarchaeal assemblages within mesophilic soil ecosystems as revealed by PCR-single-stranded conformation polymorphism profiling. *Appl. Environ. Microbiol.* **70**, 1811–1820 (2004).

96. Sliwinski, M. K. & Goodman, R. M. Comparison of Crenarchaeal consortia inhabiting the rhizosphere of diverse terrestrial plants with those in bulk soil in native environments. *Appl. Environ. Microbiol.* **70**, 1821–1826 (2004).

97. Riemann, L. & Middelboe, M. Stability of bacterial and viral community compositions in Danish coastal waters as depicted by DNA fingerprinting techniques. *Aquat. Microb. Ecol.* **27**, 219–232 (2002).

98. Pinhassi, J. *et al.* Spatial variability in bacterioplankton community composition at the Skagerrak–Kattegat Front. *Mar. Ecol. Prog. Ser.* **255**, 1–13 (2003).

99. Troussellier, M. *et al.* Bacterial activity and genetic richness along an estuarine gradient (Rhone River plume, France). *Aquat. Microb. Ecol.* **28**, 13–24 (2002).

100. Casamayor, E. O. *et al.* Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environ. Microbiol.* **4**, 338–348 (2002).

101. McArthur, J. V., Kovacic, D. A. & Smith, M. H. Genetic diversity in natural populations of a soil bacterium across a landscape gradient. *Proc. Natl Acad. Sci. USA* **85**, 9621–9624 (1988).

102. Buckley, D. H. & Schmidt, T. M. Diversity and dynamics of microbial communities in soils from agro-ecosystems. *Environ. Microbiol.* **5**, 441–452 (2003).

103. Staddon, W. J., Trevors, J. T., Duchesne, L. C. & Colombo, C. A. Soil microbial diversity and community structure across a climatic gradient in western Canada. *Biodivers. Conserv.* **7**, 1081–1092 (1998).

104. Franklin, R. B., Taylor, D. R. & Mills, A. L. The distribution of microbial communities in anaerobic and aerobic zones of a shallow coastal plain aquifer. *Microb. Ecol.* **38**, 377–386 (1999).

105. Franklin, R. B. & Mills, A. L. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiol. Ecol.* **44**, 335–346 (2003).

106. Martiny, A. C., Jorgensen, T. M., Albrechtsen, H. J., Arvin, E. & Molin, S. Long-term succession of structure and diversity of a biofilm formed in a model drinking water distribution system. *Appl. Environ. Microbiol.* **69**, 6899–6907 (2003).

107. Franklin, R. B., Blum, L. K., McComb, A. C. & Mills, A. L. A geostatistical analysis of small-scale spatial variability in bacterial abundance and community structure in salt marsh creek bank sediments. *FEMS Microbiol. Ecol.* **42**, 71–80 (2002).

108. Rohwer, F., Seguritan, V., Azam, F. & Knowlton, N. Diversity and distribution of coral-associated bacteria. *Mar. Ecol. Prog. Ser.* **243**, 1–10 (2002).

109. Yannarell, A. C. & Triplett, E. W. Geographic and environmental sources of variation in lake bacterial community composition. *Appl. Environ. Microbiol.* **71**, 227–239 (2005).

110. 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).

Competing interests statement
The authors declare no competing financial interests.