

Microbial community structure and its functional implications

Jed A. Fuhrman¹

Marine microbial communities are engines of globally important processes, such as the marine carbon, nitrogen and sulphur cycles. Recent data on the structures of these communities show that they adhere to universal biological rules. Co-occurrence patterns can help define species identities, and systems-biology tools are revealing networks of interacting microorganisms. Some microbial systems are found to change predictably, helping us to anticipate how microbial communities and their activities will shift in a changing world.

Microorganisms, by which I mean Bacteria, Archaea, viruses, protists and fungi, are vital to the function of all ecosystems. This is largely because they exist in enormous numbers (there are roughly 5×10^{30} bacteria alone worldwide)¹ and so have immense cumulative mass and activity¹. They are also probably more diverse than any other organisms, so it is easy to see why the structure of microbial communities, that is, the different kinds of organisms and their abundances, is so important to the way in which ecosystems function. But even with modern tools, it is not easy to determine microbial community structure and map its variations in space and time; we have only recently begun to learn about the scales of variation (Box 1). Understanding ecosystem function, and predicting Earth's response to global changes such as warming and ocean acidification, calls for much better knowledge than we have today about microbial processes and interactions.

In the past couple of decades, the use of genome sequences and related approaches^{2–6} has overcome the need for cultivation to characterize and identify microorganisms in nature (Box 2). It still seems almost hopeless to sort out the identities and interrelationships among the trillions of microorganisms in a cubic metre of sea water, let alone a few hectares of ocean, but high-throughput sequencing and whole-community fingerprinting techniques have enabled researchers to make considerable progress (Box 2). This review summarizes our knowledge of microbial community structure, with a focus on planktonic marine bacteria, and discusses what we can learn about microbial systems and their functions from this information (Box 3).

General distributions and the 'rare biosphere'

The classic dictum about microbial distribution patterns, "everything is everywhere but the environment selects", is attributed to Lourens Baas Becking⁷. This concept has been thoughtfully reviewed in recent years^{8–12}, often with particular attention to the long tail of the species abundance curve, which shows that large numbers of individually rare species are found in most ecosystems (Fig. 1). This concept reflects the fact that current distributions of organisms are the result of historical factors, including dispersion by wind, water and animals, and adaptations to local conditions that change over space and time. The 'everything is everywhere' part alludes to the remarkable dispersal potential of microorganisms. It has been claimed, largely on the basis of studies of some morphologically defined protistan species, that for species comprising organisms that are smaller than 1 mm, global diversity is relatively low and organisms are cosmopolitan, essentially lacking biogeographical

variations¹³. This is presumably owing to their high population sizes and easy dispersion. It is impossible to disprove this assertion, because we cannot prove the complete absence of an organism. Nevertheless, in practice and to the extent that most measurements allow (Box 2), numerous studies suggest that most microorganisms do not seem to be cosmopolitan, even within a given habitat type, and discernible biogeographical patterns are typical^{9,10,14}. The idea that 'the environment selects' indicates that only those organisms capable of activity and growth in a particular environment will increase in number. There is a fuzzy boundary between 'common' and 'rare' organisms, often described as somewhere in the range of 0.1–1% of the microbial community. It has been argued that the rare ones do not ordinarily affect most major biogeochemical processes, such as respiration and the processing of nutrients⁸. Studies indicate that in marine plankton, common organisms typically carry out most of the activity^{15,16}. Cultivation techniques often yield bacteria that are very rare in whole-community cloning studies, such as *Vibrio* spp. By contrast, most molecular studies tend to find primarily the common organisms⁸, and characterizing the rare ones with molecular tools typically requires a focused study, such as high-throughput tag sequencing^{17,18}.

Among the more common, and presumably more active, microorganisms, molecular survey data can be used to show the extent to which particular organisms are cosmopolitan, widespread or endemic within a given habitat type and can relate distributions to other properties. One study¹⁹ of marine-plankton clone libraries collected samples from nine widespread locations, with 263–702 clones each, and coverage estimates in the range 45–94%. Of 582 unique operational taxonomic units (OTUs), defined in that study as having at least 97% 16S ribosomal RNA similarity, 69% were found only at a single location (endemic), 17% were at two locations, 6% at three locations, and only 0.4% were cosmopolitan (found at all nine locations). The proportion of endemic ones was similar, irrespective of community size. The more widespread OTUs also tended to be the most abundant at individual locations. The endemic OTUs tended to be individually rare; 92% of them had only 1–2 clones, but four endemic OTUs were relatively abundant, with 7–24 clones at their respective locations. A simplistic model of these data is that an organism abundant in one place is likely to be detectable and may be abundant elsewhere, whereas one that is rare but detectable (0.1–1% by the approach used in this study) in one location is likely to be below the detection limit in other locations. Many of these rare ones are presumably present elsewhere but are too rare to detect easily, or they may become abundant at some other place or time. The apparent relationship between abundance and range size is one of several

¹Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA.

instances in which microbial patterns parallel those of larger organisms¹⁴ (discussed further in the section 'Universal patterns').

The apparently low contribution of rare organisms to most biogeochemical processes does not mean they are unimportant. Yes, in a given environment, most may simply be passing through and may die without ever finding a suitable niche. But sometimes the highly diverse members of the 'rare biosphere'¹⁷ can have clear ecological significance. The most obvious influence is by acting like a seed bank⁸. Organisms that may be ideally adapted to conditions in another time or place could eventually thrive by just dispersing and waiting. Many microorganisms that are rare in a given plankton sample might be in the process of dispersion among adjacent niches in which they may be extremely abundant; for example, in suspended aggregates of 'marine snow', in the guts or on the surfaces of animals (for instance, pathogens such as *Vibrio vulnificus*), or in sediments. Many that are rare at a given time probably thrive during a different season or in occasional phenomena such as El Niño events. Marine bacteria that are rare in one season can be abundant in another. For example, in a four-year time-series study²⁰, a variety of taxa were undetectable in some months (<0.1% of the total), but then made up several per cent of the community in other months. In some cases, the seed-bank aspect may apply not to whole organisms but just to their genes¹⁷. A rare individual may have some genes that when transferred to another organism will create a recombinant that is better adapted to a particular habitat than either parent. The seed bank can function as a valuable insurance as global and local conditions change through natural or anthropogenic causes.

Even some 'chronically rare' organisms may have global biogeochemical significance. There is no mechanism to prevent the long-term persistence of widespread but rare organisms, and it stands to reason that rare organisms are probably growing slowly (otherwise they would become common). It has been argued that being rare and slow-growing helps them avoid major mortality processes such as viral infection and predation⁸. The viral infection rate is directly dependent on the abundance of hosts. Predation by protists generally removes the larger organisms that tend to be growing faster, so rare ones are presumably smaller and grazed on less⁸. Chance extinction as a result of stochastic events is reduced by enormous microbial population sizes; planktonic bacteria, representing only 1 in 10,000 cells, still have 10¹⁴ individuals per cubic kilometre. In addition, rare organisms can collectively affect globally important biogeochemical processes. For example,

marine nitrogen fixation — a significant input of nitrogen into the global biosphere — was thought to be performed primarily by *Trichodesmium*, a globally rare but locally (and episodically) abundant colonial cyanobacterium²¹, or by symbiotic cyanobacteria in larger phytoplankton²². But recent evidence suggests that chronically rare cells, which represent about 1 in 10,000 of the total marine bacteria, may at times cumulatively fix more nitrogen than the larger organisms²³. Perhaps other rare organisms with unique capabilities are important for processes such as slowly breaking down recalcitrant dissolved organic compounds in sea water.

Difficulties evaluating dispersion

The distribution of organisms is controlled by dispersion and adaptations to environmental heterogeneity¹⁴, so understanding dispersion is important. We can demonstrate that dispersion has happened by finding the same kind of organism in different places. But the OTUs used to identify 'sameness' in microbial studies are not equivalent to species in larger organisms. Frederick Cohan²⁴ suggested conservatively that what is called a single species in bacteria might be as broad as a genus in the macrobiota, and James Staley²⁵ made the point that the formal definition of a bacterial species would classify all the primates from humans to lemurs as a single species. There is not a precise 16S rRNA 'molecular clock', but a 1% change in bacterial 16S sequence is estimated to take millions of years^{26,27}. This is important in evaluating the role of dispersion, because what we call the 'same' microorganism, on the basis of membership in an OTU or even an accepted species, may be so varied as to be irrelevant for dispersion studies. So we should probably be considering 'microdiversity'^{28–30} when looking for dispersion as a mechanism to explain contemporary distributions. One would expect that only individuals that are essentially identical in slowly evolving phylogenetic markers such as 16S rRNA would be suitable as evidence for recent dispersion.

Universal patterns

Do microorganisms follow well-established patterns previously observed in larger organisms, implying the presence of 'universal' processes or profound unifying rules that apply to all life? Or is there something fundamentally different in character about these smallest organisms that allows us to distinguish between rules that are truly universal and those that are not? Growing evidence suggests that microorganisms do follow some classic ecological patterns, bolstering the likelihood of successfully applying established ecological theory³¹. Examples are latitudinal gradients in diversity, taxa–area relationships and community assembly 'rules'.

Latitudinal gradients

One of the oldest observed patterns of animal and plant diversity, first reported in the early nineteenth century, is the tendency of lower latitudes to have more species than higher latitudes^{32–34}. There are several possible causes that are not mutually exclusive and continue to be debated^{32–34}. One hypothesis is that on land, generally higher productivity in lower latitudes provides more resources that can be split into more niches. Another hypothesis is that higher temperatures in low latitudes increase the metabolic rate and make biological processes, including speciation, occur faster. Does this latitudinal gradient apply to microorganisms, and if so, can the particular pattern help us choose between the hypotheses? Some reports suggest that there is little or no microbial gradient¹³. Meta-analysis of data on a wide variety of species (from protists to megafauna)³⁴ or exclusively on marine species³⁵ indicated that the strength and slope of the latitudinal richness gradient is reduced as organism size decreases, which could be extrapolated to suggest that there is little or no gradient for bacteria-sized organisms. A fingerprinting study of soil bacteria found a moderate range of richness, best predicted by soil pH, with no discernible latitudinal gradient³⁶; acidic soils had much lower diversity than neutral or alkaline soils, irrespective of latitude and temperature. In this case, it seems that an overriding influence of soil pH combined with high natural variability in soil habitats (and differing sampling effort¹⁴) could have masked other, more subtle, influences on diversity.

In contrast to these results, two recent reports found a bacterial latitudinal richness gradient in marine plankton. One was a study of

Box 1 | Scales of variation

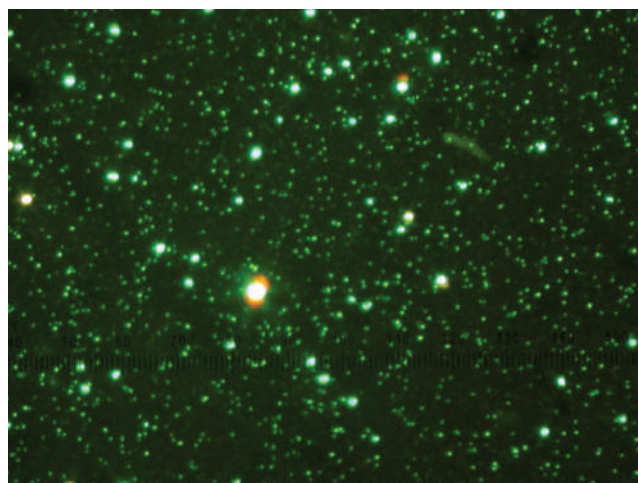
Changes in the community structure in space and time are very informative. For a start, they show us what scales a particular sample represents. This is crucial for extrapolating from individual samples to the world at large, a common exercise that is often not well informed. For example, samples collected for the *Sorcerer II* Global Ocean Sampling (GOS) Expedition⁷⁴ were typically tens of litres of sea water from near the surface. What region of the ocean (horizontal and vertical), and over what timescale, might the microorganisms in such a sample represent quantitatively? Recent community fingerprinting studies with similar samples of marine plankton have found timescales for significant community changes to be typically of the order of days and weeks^{51,75}. In terms of size scales, patches of coherent communities at a given depth horizon tend to be on the scale of kilometres or tens of kilometres^{51,75}. In the vertical direction, there can be significant changes over metres or tens of metres, or even over millimetres at the immediate sea-surface microlayer⁷⁶. This suggests a given GOS sample is likely to represent a range of about a week temporally, a few kilometres horizontally and a few metres or tens of metres vertically. But beyond this practical application, changes in community structure help us to understand factors that control communities. For example, the horizontal 10-km scale of variability resembles that of mesoscale ocean eddies of factors such as chlorophyll, as observed by satellite⁵¹, suggesting that the physicochemical and biological factors that structure the microbial community composition may relate to, or even be the same as, those that control phytoplankton abundance.

Box 2 | Measuring and comparing community composition

Most contemporary studies of microbial community composition are DNA-based, avoiding cultivation, which misses most organisms⁷⁷. DNA extracted from the microbial community (the 'metagenome') can be analysed in several ways:

- (1) by cloning and sequencing phylogenetically informative genes, such as those for 16S ribosomal RNA;
- (2) by analysing all the genes studied, either by random cloning into large or small insert libraries and sequencing those, or by clone-free methods such as pyrosequencing;
- (3) by using high-throughput sequence analysis of phylogenetically informative short sequence tags¹⁷;
- (4) by using whole-community fingerprinting methods such as terminal restriction fragment-length polymorphism (TRFLP), denaturing gradient gel electrophoresis (DGGE) or automated ribosomal intergenic spacer analysis (ARISA). These less costly fingerprinting methods yield data (such as gene fragment sizes for TRFLP and ARISA) on specific components of the community, but provide less information about individuals and usually allow the putative identification of components by comparison with sequences. The organisms from which the gene sequences or fingerprints were derived are categorized, or 'binned', into operational taxonomic units (OTUs) on the basis of a variety of criteria that can vary with the method, such as sequence similarity or fragment size. Some studies do not use extracted DNA but instead involve tagging and counting particular organisms (one OTU at a time) by fluorescence *in situ* hybridization. With all these methods, quantitative interpretation of the results must consider possible biases, which can occur at several steps from collection to binning⁷⁷. Some approaches are reasonably quantitative for certain OTUs, such as abundance estimates of *Prochlorococcus* by ARISA²⁰. Even imperfect methods, like some used in ecology (for example, fogging a tree with insecticide to get 'all' the insects to fall into a net), have been used to develop fundamental theories.

Comparing community composition between samples is challenging when the assay covers only a minority of each community, as occurs in highly diverse samples analysed by limited numbers of sequences. High sequence coverage helps and, for moderately diverse samples such as marine planktonic bacteria, a few hundred 16S rRNA clones are estimated to have about 50–90% OTU coverage¹⁹. Alternatively, whole-community fingerprinting, by methods such as ARISA, which can show essentially all the bacterial OTUs making up more than 0.1% of the community⁷², allows



cost-efficient comparison of multiple samples^{36,72,75}, as long as one can ignore organisms that each represent less than 0.1% of the population. Similarly, comparing OTU richness (the total number of measured OTUs) or species lists between samples is challenging because each protocol catches a different proportion of the rare OTUs (no protocol can catch them all). They therefore typically compare lists with different levels of completeness, even with a massive sequencing effort¹⁸. The fairest comparisons use standardized measures of composition or richness, such as by fingerprinting³⁷, or a uniform and high-coverage technique to generate sequences. This problem is well known in ecology, and statistical approaches have been used^{17,72} and continue to be developed^{17,73,78,79} to estimate richness and compare diversity while taking into account the varying amounts of undetected organisms.

The image is an epifluorescence micrograph of marine viruses (smallest green dots), bacteria and/or archaeal organisms (medium-size green dots) and pigmented protists (larger green dots with red patches), representing a small part of a microlitre of sea water from the northeast Pacific Ocean off California. For scale, the bacteria average about 0.5 μm in diameter. The non-descript nature of these organisms means that we need molecular techniques to characterize the community composition.

nine samples using 16S rRNA clone libraries¹⁹, and the other was a whole-community fingerprinting study³⁷ that examined 103 samples from near the sea surface at 56 locations worldwide. In the latter study, the measured richness values, when plotted against latitude, fell within triangular constraint envelopes whereby high latitudes had low richness, but lower latitudes could have high or low richness. This suggests that latitude sets an upper limit on richness but that other factors can sometimes reduce it. In both studies, richness was correlated with temperature, but not with indices of productivity such as chlorophyll or annual primary production, consistent with the theory that the metabolic rate, affected by temperature, may strongly influence the pace of complex ecological processes such as speciation³⁷. Why is this different from soils? Perhaps by studying near-surface plankton in sea water, a habitat much more uniform than soils, many chemical and physical factors were held relatively constant, allowing the detection of an effect related to latitude despite the presence of other important factors.

Taxa-area relationships

Another widely studied ecological pattern is the relationship between the number of observed species (S) and the sampled area (A), commonly assumed to have a power-law relationship, $S \propto A^z$. Values for z are determined empirically, and for animals and plants are typically 0.1–0.3 for contiguous habitats and 0.25–0.35 for islands³³. Studies of microorganisms (including marine planktonic diatoms and salt-marsh bacteria, but not planktonic bacteria) have found a range of z values, often lower than 0.1, but some recent studies have reported

z values that overlap the canonical ranges, especially for 'island-like' situations^{12,14,31,38,39}. The broader definition of a microbial (as opposed to animal or plant) species lowers z significantly³⁸, and z is also affected by the difficulty in comprehensively sampling microorganisms^{14,39}. Nevertheless, the overlap in values suggests that relationships for microorganisms can be similar quantitatively to those of larger organisms, indicating that this may be a universal rule for all domains of life.

Related to taxa–area relationships are distance–decay relationships, which show community differences with increasing distance¹⁴. For deep-sea bacteria, fingerprinting shows a significant and steady decline in community similarity over a 3,500-km distance at a depth of 3,000 m (in the Pacific Ocean), and over a 1,000-km distance at a depth of 1,000 m (Atlantic Ocean near the Amazon plume), with much scatter at 1,000 m (Pacific) and 500 m (Atlantic)⁴⁰. These results dispel the notion that the deep sea is uniform in microbial communities, and show that different depths and locations have different relationships, with considerable patchiness. One hypothesis⁴⁰ is that differences in patterns may relate to shallower and mid-water depths being most influenced by the patchy 'raining' of detritus that provides food to the deep sea, with most organic carbon being consumed before it reaches the lower depths.

Community assembly 'rules'

Co-occurrence patterns of organisms — examining which organisms sometimes or never occur together — have been used to reveal community assembly rules^{41,42}. Several ecological processes potentially contribute to these nonrandom co-occurrence patterns, including competition,

Box 3 | Applications of microbial community-structure studies

Microbial community structure can be used to address several timely issues.

- Finding basic patterns in microbial distributions, such as the extent of endemism and ubiquity, and learning how to evaluate these given the enormous 'rare biosphere' and difficulties characterizing dispersion
- Searching for general patterns that might apply to all life and support the concept of 'universal' rules that broadly control biology, thereby improving our theories and leading to predictions
- Using co-occurrence patterns to help define potentially interacting organisms and interaction networks in highly complex systems
- Using co-occurrence of organisms and correlations to environmental parameters to help delineate microbial species
- Examining seasonally or annually repeating patterns, also leading to predictions
- Evaluating the roles of viruses
- Linking community structure or particular genes to particular functions

habitat filtering, historical effects and neutral processes⁴³. One study⁴⁴ compared co-occurrence in more than 100 microbial (Bacteria, Archaea, protists and fungi) data sets with a meta-analysis of almost 100 macro-organism data sets. The microbial assemblages had nonrandom patterns of co-occurrence broadly similar to those in macroorganisms. The authors concluded that some co-occurrence patterns may be general characteristics of all domains of life. The extent of co-occurrence in microbial communities did not vary between broad taxonomic groups or habitat types. There were variations from using different methods to survey microbial communities (such as clone libraries or fingerprinting), and taxonomic resolution was also a factor. The authors also noted that undersampling of microorganisms may underestimate the extent of community structure, so microorganisms might have more highly structured communities than macroorganisms.

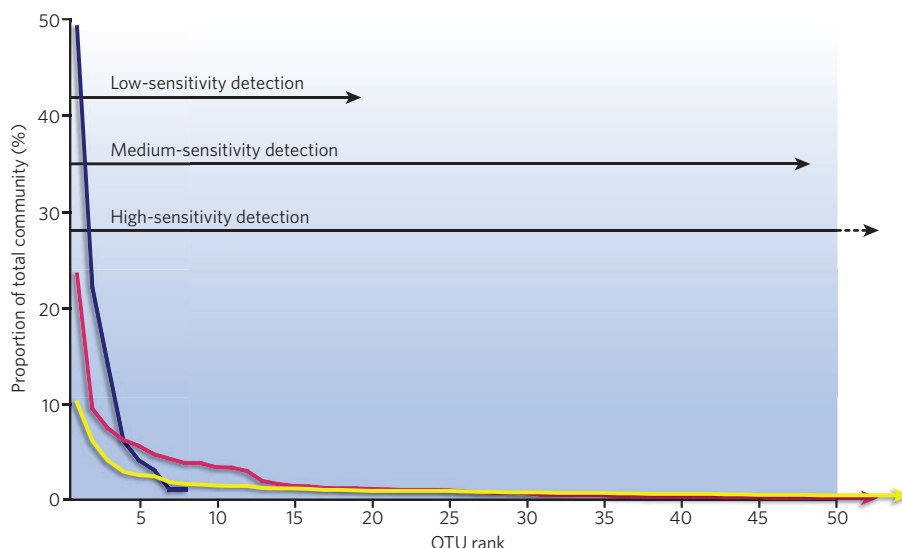


Figure 1 | Rank abundance relationships for bacterial operational taxonomic units. Different communities vary immensely in species richness. The examples here represent an extremely acidic habitat with low diversity (blue line)⁷¹, marine plankton with moderate diversity (red line)⁷² and sediment with high diversity (yellow line)⁷³. All environments are dominated by just a few operational taxonomic units (OTUs), but have a long tail in the distribution (rare taxa) that can number in the tens of thousands, off the figure to the right. Low-sensitivity methods such as 100-clone libraries show only the most abundant taxa, but that may be sufficient for extremely low-diversity samples. Methods with moderate sensitivity such as 1,000-clone libraries or ARISA (automated ribosomal intergenic spacer analysis) fingerprinting show a larger fraction of the community, but still can miss many rare taxa. High-throughput methods such as tag sequencing¹⁷ can detect thousands of taxa. It is a practical impossibility to detect all of the taxa. Different studies require knowledge of more or less of these rare taxa to support their conclusions, so it is important to select the appropriate method.

Interactions and networks

Microbial communities, as part of natural ecosystems, are inherently complex. The traditional tools of microbiology, such as pure cultures and genetic studies, tend to provide a reductionist view, studying each organism in isolation. However, a reductionist approach is not well suited for learning about interactions and emergent properties of communities. Holistic approaches, which study natural habitats directly, can yield complementary data to help deduce the interactions. Co-occurrence patterns, as described above, also show how particular organisms in a system occur together and vary with environmental parameters. These patterns show important details of a particular community structure and can be represented as mathematical interaction diagrams or networks. If environmental conditions are included in the co-occurrence patterns, the results indicate which conditions the co-occurring assemblages of organisms prefer or avoid. The inclusion of biogeochemical rate measurements may link groups of organisms with particular functions. These kinds of result allow us to examine the potential interactions between organisms and aspects of the niches of microorganisms within extremely complex and dynamic natural communities.

Such microbial network studies are at an early stage and, given the high natural variability, they require a large amount of data if the findings are to be statistically significant. They also require suitable analytical tools for the evaluation and sorting of enormous numbers of potential interactions. Some of the initial studies^{45,46} used data from the San Pedro Ocean Time-Series (SPOT) microbial observatory site (southern California). Quansong Ruan and colleagues⁴⁵ developed 'local similarity analysis' to evaluate contemporaneous and time-lagged correlations between parameters, and discerned a portion of a network using this approach. Time-lagged relationships (for example between predators and prey) are common in ecology. Joshua Steele and I⁴⁶ extended this analysis using tools from systems biology to visualize the results as networks centred on members of particular groups of organisms connected to their 'nearest neighbours', which are organisms or parameters that

directly correlate positively or negatively. An example of such a network, centred on members of the ubiquitous SAR11 cluster, shows ten different members of this cluster, each with distinctive combinations of positive and negative correlations to other bacteria or environmental parameters (Fig. 2). Previous work had shown the seasonality of a few broadly defined SAR11 types⁴⁷, but this network analysis provides much more detailed distinctions. Interestingly, only two pairs of SAR11 subtypes correlated to each other, and the higher correlated pair were members of different subclades. This demonstrates that ecological relatedness may not follow phylogenetic relatedness, even in a narrow phylogenetic group.

In these networks, the 'interactions' are mathematical relationships (correlations or anticorrelations) that need further investigation to distinguish direct interactions from indirect ones. This is an important distinction, partly because classically studied ecological networks, on which there is an extensive body of work⁴⁸, consider only direct interactions. Positive correlations in these mathematically derived networks could be common preferred conditions or perhaps cooperative activities such as cross-feeding. Similarly, negative correlations may represent opposite seasonality, competition for limited resources or perhaps active negative interactions such as targeted allelopathy⁴⁹ or predator-prey relationships⁵⁰. The ability of local similarity analysis to see lagged relationships, showing when one parameter changes in

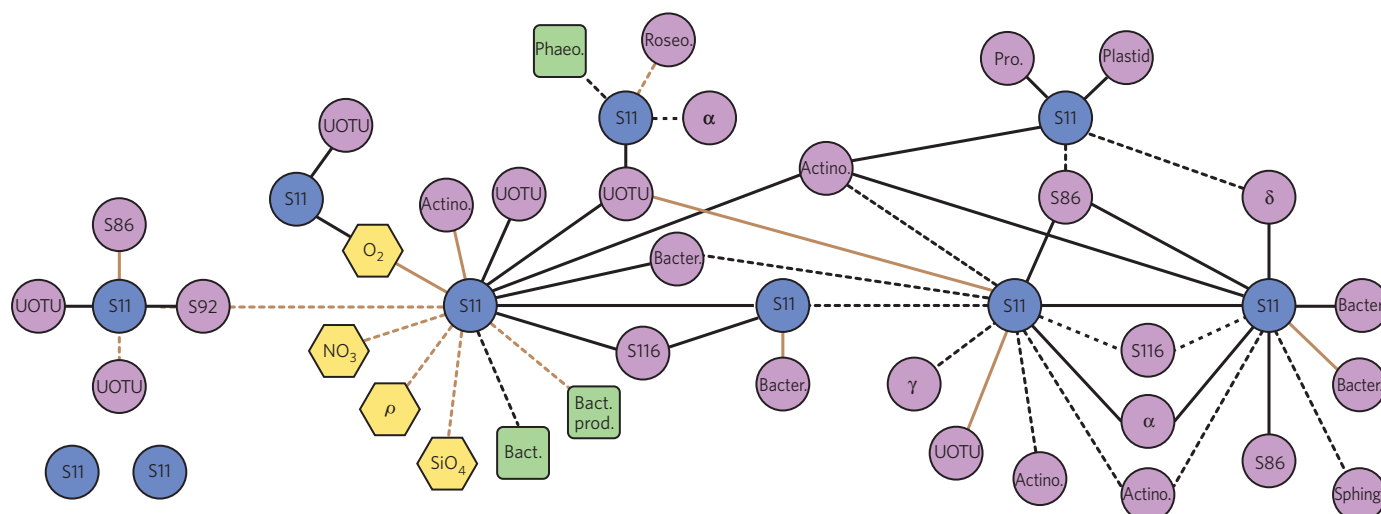


Figure 2 | Association networks. Systems-analysis tools can create association networks from field data on distributions, and these networks can have a time component if the measurements are a time series. This example from a 4.5-year time series off the California coast shows only the SAR11 bacteria and the organisms or parameters that directly correlate to them. There are ten discernible SAR11 subtypes (blue circles), but none of them shares the same combinations of correlations to abiotic (yellow hexagons) and biotic (green squares) environmental parameters or other individually distinguishable taxa (purple circles). Most of the SAR11 subtypes correlate to multiple bacterial uncharacterized operational taxonomic units (UOTUs) only, and few show correlations to environmental

parameters directly. These relationships suggest that the ten SAR11 subtypes are candidates for different species, and that the differences between them are shown most clearly by their correlations to other bacteria and not by the relationships with environmental parameters. Solid lines indicate positive correlations. Dashed lines indicate negative correlations. Correlations with a 1-month time lag are indicated by brown lines. Data taken from ref. 46. α, α-proteobacteria; γ, γ-proteobacteria; δ, δ-proteobacteria; ρ, density; Actino., Actinobacteria; Bact., total bacteria; Bacter., Bacteroidetes; Bact. prod., bacterial production; NO₃, nitrate; O₂, oxygen; Phaeo., phaeopigments; Pro, *Prochlorococcus*; Roseo., *Roseobacter*; S, SAR; SiO₄, silicate; Sphing, *Sphingobacterium*.

advance of another, may help to sort out cause and effect.

To capture the important interactions that regulate system functions, network analysis should include all the important parameters in that system; the early networks, as shown in Fig. 2 with bacteria and some environmental parameters, show only the tip of the iceberg. Because microbial communities clearly include many interactions with protists, viruses and metazoans⁵¹, all these organisms should ultimately be included in the analysis, otherwise important controlling factors will be missed. Considerable progress has been made in the molecular genetic characterization of marine protists^{52–54} and viruses^{55–57} (see also page 207). An initial analysis extending the interactions shown in Fig. 2 with Archaea and protists included shows that each bacterial OTU tends to have specific and multiple interactions with protist OTUs; sometimes the protist community seems more important than physicochemical parameters in determining the bacterial community structures (J. A. Steele, P. D. Countway, J. M. Beman, L. Xia, J. Huang, P. D. Vigil, F. Sun, D. A. Caron, J. A. F., unpublished work). Network analysis can also be integrated with 'omics' studies (see page 200) for a more complete picture of organism functions in an environmental context. Holistic studies show potential interactions, and focused omics studies relate functions to organisms, illuminating the mechanism of the interactions.

Defining microbial species

Much debate has centred on the definition of microbial species, for theoretical and practical reasons^{24,58,59}. What we call a species, and why, influences how we think about, study and understand organisms. It is relevant here to discuss how community structure can help evaluate species identities. Recent reports have demonstrated the value of characterizing the microenvironmental (ecological) preferences of organisms, such as season and size fraction for marine *Vibrio* spp.⁶⁰ or canyon microhabitat for *Bacillus* clades⁵⁹, to help verify the distinctiveness of closely related, genetically defined clusters that resemble species. I suggest that extending the analysis to include the co-occurrence of organisms can also help define ecological niches and lead to better ways of characterizing ecological species or microbial ecotypes. In other words,

we should consider the biological environments of microorganisms, as well as their physicochemical environments, as important aspects of their niches. This is because many organisms, particularly microorganisms, live in microenvironments that are probably defined at least as much by the organisms around them as by physicochemical characteristics. Organisms create many aspects of the environment themselves (including subtle ones such as dilute nutrients), and their presence also provides a sort of monitoring and long-term historical integration of environmental conditions that are hard to measure directly, yet that may relate to the essence of a niche. An example would be to consider the ten SAR11 types in Fig. 2 as candidates for different ecological species or ecotypes. Obviously, more information than these correlations would be needed to confirm such distinctions (not easy in the case of organisms in their natural habitats), but networks provide a starting point and can hasten the development of descriptive or predictive models.

Predictable patterns of change

Multi-year monitoring by fingerprinting of SPOT samples reveals repeating patterns of bacterial community composition that are predictable not only on the basis of seasons (the month of occurrence), but also on the basis of environmental conditions such as temperature, salinity, chlorophyll and nutrient abundance, and viral and bacterial abundance⁶¹. Here, predictability refers to the statistical analysis of past data, but it suggests the possibility of forecasting changes in microbial communities. By expanding such approaches, we can start to consider how global change might alter the microbial landscape locally or globally.

The predictability of community composition from environmental parameters has some profound implications. It implies the presence of well-defined, probably narrow, niches for the predictable organisms. It also implies the related conclusion that the combinations of functional properties in each of the predictable organisms is unique (within these environments, at least), because redundancy or interchangeability between organisms would lead to random replacements that would interfere with the predictions. This non-redundancy is consistent with classical competitive exclusion or niche partitioning leading to one

organism in each niche. Predictability also implies a stable and consistent phenotype associated with the ARISA-defined genotype used to characterize members of the communities (the length of the spacer between 16S and 23S rRNA genes). This has implications for our interpretation of genetic data in general, as we discuss later.

Viral effects

Viruses are the most abundant biological entities in the sea, typically being an order of magnitude more abundant than bacteria. They are crucial components of marine systems, having major effects on ecological, biogeochemical and evolutionary processes^{55,62–64} (see also page 207). They have left their genetic tracks widely in marine microbial genomes, for example in *Prochlorococcus*⁶⁵, and have short-term and long-term effects on the community structure of bacteria and protists⁵¹. Of particular relevance is the ‘kill the winner’ hypothesis⁶⁶, which posits that because virus infection is generally host specific and density dependent, abundant organisms are most susceptible to epidemic infection, and infection would reduce their abundance, leading to a succession of dominant organisms over time as one after another is attacked^{162,66,67}. This is the prediction based on theory, but when we look at field data, the dominant OTUs often stay dominant for extended periods, certainly longer than this theory predicts⁵¹. One possible explanation is that OTUs are made up of multiple ‘strains’ with different virus sensitivities, and a closer examination would reveal a succession of different dominant strains. OTU definitions are typically based on slowly evolving genes that are generally not transferred (such as 16S rRNA), and virus resistance is often determined by more rapidly evolving or more easily transferred genes.

Linking community structure to genes and functions

Does the presence of an organism, an OTU or a functional gene tell us about a particular biogeochemical function? This is a fundamental question about the applicability of genetic studies to investigations of ecosystem function. Transcriptomics or proteomics will help (see page 200), but such studies are costly and the data are not as widely available as genetic information. The question arises because many genes are expressed only sometimes by the organism, such as the nitrogenase gene during nitrogen fixation²². In addition, different members of closely related groups of organisms — even well-defined species such as *Escherichia coli* — can have widely different genomes²⁴, so finding the species or OTU by a phylogenetic marker may not necessarily mean that other particular genes accompany it.

Core and flexible genomes

Intraspecific genomic consistency brings up the concept of ‘core’ and ‘flexible’ genomes of organism clusters, for example in the abundant marine cyanobacterium *Prochlorococcus*. Extensive study of numerous isolates of *Prochlorococcus*⁶⁵ has found that the genomes share more than 1,200 core genes (the ‘core genome’) that tend to perform functions of central housekeeping, such as DNA and protein synthesis. These core genes are thought to code for a functional cell, and phylogenetic trees constructed from them are generally congruent, implying little transfer between lineages⁶⁵. These organisms also have many genes present only in some isolates. The 12 sequenced *Prochlorococcus* isolates collectively have about 4,000 flexible genes (there is usually a similar number of flexible and core genes in a given cell). Collectively called the ‘flexible genome’, these genes typically encode accessory functions such as the transport or use of certain non-essential nutrients, protection against damage from high light levels, and certain cell-surface modifications that probably relate to phage or predator resistance, but many have no known function. These flexible genes apparently control important aspects of the niche adaptations that define where the organisms can thrive. They are gained, lost or transferred much more than the core genes. The topic of core and flexible genomes is highly relevant for the interpretation of how community structure relates to function, because we use sequences or markers from the core genome to identify organisms, but genes in the flexible genome often define many of the

functional characteristics.

Is the relationship between the core and accessory genome so variable that we cannot predict relevant environmental functions from core genome data such as 16S rRNA genes? It is too early to say whether the divergence between core and flexible genomes is a serious problem for interpreting community structure, but so far the data suggest that it is not a great problem. In *Prochlorococcus*, many genes coding for obvious niche-defining features such as optimal light levels and use of common nutrients seem to be shared among the most closely related strains⁶⁵, which is encouraging. Empirical data on distributions of six 16S-rRNA-based *Prochlorococcus* ecotypes, over wide areas of the ocean, suggested sensible distribution patterns that were consistent with stable phenotypes⁶⁸. The 16S rRNA sequence differences between ecotypes were small, less than 3%, so OTUs were defined narrowly. The data on repeating patterns and predictability of bacterial community composition at the SPOT site⁶¹ suggest that the relationship between phylogenetic markers and niche-defining phenotypes is usually fairly predictable. These field results imply that the marker genes (ARISA 16S–23S rRNA spacer length) were consistently associated with the part of the phenotype that made the organism predictable (its niche). They suggest that a particular core genome mostly has a consistent flexible genome. In the assemblage as a whole and over time, there seems to be a similar and consistent set of accessory genes associated with each core genome; that is, if there are multiple accessory genomes and phenotypes associated with a particular core genome, the proportions stay fairly constant. This further implies general stability of the community collective genome over ecological timescales. Studies currently under way will show whether these early encouraging results hold up.

Linking functions with genes

One goal of metagenomics (see pages 200 and 207) is to link particular functions to particular organisms. This requires the annotation of gene sequences with their functions, which can be a challenge, as shown by the example of proteorhodopsin, one of the most exciting early discoveries of metagenomics⁶⁹. Initial analysis showed that this protein can act like a light-driven proton pump, allowing the cell to generate usable energy from light (phototrophy)⁶⁹. But more recent results, including the lack of any clear growth benefit from light in most cultured organisms that have the gene, question whether phototrophy commonly occurs in the many organisms with this gene and whether the gene may have other functions as well⁷⁰. So it may not be straightforward to interpret the function of a gene found in the environment, even when studied extensively. Linking genes with functions is probably best done by the integration of metagenomics, cultivation and field measurements.

Future prospects

New low-cost, high-throughput sequencing can greatly advance the analysis of marine microbial community structure, especially for measurements spread over space and time. Difficulties will remain, however, such as discovering the distributions of rare organisms. But this need not stop us following, modelling and eventually predicting the distributions of the majority of microorganisms and their activities, a critical aspect of understanding biogeochemical cycles in our changing world. ■

- Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
 - Pace, N. R., Stahl, D. A., Lane, D. L. & Olsen, G. J. The analysis of natural microbial populations by rRNA sequences. *Adv. Microb. Ecol.* **9**, 1–55 (1986).
 - Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
 - Angly, F. E. et al. The marine viromes of four oceanic regions. *PLoS Biol.* **4**, e368 (2006).
 - Massana, R. & Pedrós-Alió, C. Unveiling new microbial eukaryotes in the surface ocean. *Curr. Opin. Microbiol.* **11**, 213–218 (2008).
 - Giovannoni, S. J. & Stingl, U. Molecular diversity and ecology of microbial plankton. *Nature* **437**, 343–348 (2005).
 - Baas Becking, L. G. M. *Geobiologie of Inleiding tot de Milieukunde* (Van Stockum & Zoon, 1934).
 - Pedrós-Alió, C. Marine microbial diversity: can it be determined? *Trends Microbiol.* **14**, 257–263 (2006).
 - Martiny, J. B. H. et al. Microbial biogeography: putting microorganisms on the map. *Nature Rev. Microbiol.* **4**, 102–112 (2006).
- This review combines ideas from microbial ecologists and ‘macrobiological’ ecologists about how to include microorganisms in broader ecological theories and conceptual frameworks.

10. Foissner, W. Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozool.* **45**, 111–136 (2006).
11. Ramette, A. & Tiedje, J. M. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb. Ecol.* **53**, 197–207 (2007).
12. Fierer, N. in *Accessing Uncultivated Microorganisms: From the Environment to Organisms and Genomes and Back* (ed. Zengler, K.) 95–115 (ASM Press, 2008).
13. Fenichel, T. & Finlay, B. J. The ubiquity of small species: patterns of local and global diversity. *Bioscience* **54**, 777–784 (2004).
14. Green, J. & Bohannan, B. J. Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21**, 501–507 (2006).
15. Cottrell, M. T. & Kirchman, D. L. Contribution of major bacterial groups to bacterial biomass production (thymidine and leucine incorporation) in the Delaware estuary. *Limnol. Oceanogr.* **48**, 168–178 (2003).
16. Malmstrom, R. R., Cottrell, M. T., Elifant, H. & Kirchman, D. L. Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Appl. Environ. Microbiol.* **71**, 2979–2986 (2005).
17. 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).
18. Huber, J. A. *et al.* Microbial population structures in the deep marine biosphere. *Science* **318**, 97–100 (2007).
- This is the second of two papers describing detailed analysis of microbial community structure in marine samples studied by a high-throughput sequencing technique.**
19. Pommier, T. *et al.* Global patterns of diversity and community structure in marine bacterioplankton. *Mol. Ecol.* **16**, 867–880 (2007).
20. Brown, M. V., Schwalbach, M. S., Hewson, I. & Fuhrman, J. A. Coupling 16S-ITS rDNA clone libraries and ARISA to show marine microbial diversity: development and application to a time series. *Environ. Microbiol.* **7**, 1466–1479 (2005).
21. Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B. & Carpenter, E. J. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**, 1221–1229 (1997).
22. Zehr, J. P. & Ward, B. B. Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl. Environ. Microbiol.* **68**, 1015–1024 (2002).
23. Montoya, J. P. *et al.* High rates of N₂ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430**, 1027–1031 (2004).
24. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487 (2002).
25. Staley, J. T. Biodiversity: are microbial species threatened? *Curr. Opin. Biotechnol.* **8**, 340–345 (1997).
26. Moran, N. A., Munson, M. A., Baumann, P. & Ishikawa, H. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B* **253**, 167–171 (1993).
27. Degnan, P. H., Lazarus, A. B., Brock, C. D. & Wernegreen, J. J. Host-symbiont stability and fast evolutionary rates in an ant-bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus blochmannia*. *Syst. Biol.* **53**, 95–110 (2004).
28. Brown, M. V. & Fuhrman, J. A. Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat. Microb. Ecol.* **41**, 15–23 (2005).
29. 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).
30. Acinas, S. G. *et al.* Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**, 551–554 (2004).
31. Prosser, J. I. *et al.* The role of ecological theory in microbial ecology. *Nature Rev. Microbiol.* **5**, 384–392 (2007).
- This paper shows why the ongoing molecular revolution in microbial ecology needs to be driven by theory, and that the generality of established ecological theory must be tested using microbial systems.**
32. Pianka, E. R. Latitudinal gradients in species diversity: a review of concepts. *Am. Nat.* **100**, 33–46 (1966).
33. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
34. Hillebrand, H. On the generality of the latitudinal diversity gradient. *Am. Nat.* **163**, 192–211 (2004).
35. Hillebrand, H. Strength, slope and variability of marine latitudinal gradients. *Mar. Ecol. Prog. Ser.* **273**, 251–267 (2004).
36. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **103**, 626–631 (2006).
37. Fuhrman, J. A. *et al.* A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl Acad. Sci. USA* **105**, 7774–7778 (2008).
38. Horner-Devine, M. C., Lage, M., Hughes, J. B. & Bohannan, B. J. A taxa-area relationship for bacteria. *Nature* **432**, 750–753 (2004).
39. Woodcock, S. *et al.* Taxa-area relationships for microbes: the unsampled and the unseen. *Ecol. Lett.* **9**, 805–812 (2006).
40. Hewson, I., Steele, J. A., Capone, D. G. & Fuhrman, J. A. Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition. *Limnol. Oceanogr.* **51**, 1274–1283 (2006).
41. Gotelli, N. J. & Graves, G. R. *Null Models in Ecology* (Smithsonian Institution Press, 1996).
42. Gotelli, N. J. & McCabe, D. J. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology* **83**, 2091–2096 (2002).
43. Woodcock, S. *et al.* Neutral assembly of bacterial communities. *FEMS Microbiol. Ecol.* **62**, 171–180 (2007).
44. Horner-Devine, M. C. *et al.* A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* **88**, 1345–1353 (2007).
45. Ruan, Q. S. *et al.* Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22**, 2532–2538 (2006).
46. Fuhrman, J. A. & Steele, J. A. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. *Aquat. Microb. Ecol.* **53**, 69–81 (2008).
47. Morris, R. M. *et al.* Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. *Limnol. Oceanogr.* **50**, 1687–1696 (2005).
48. Montoya, J. M., Pimm, S. L. & Sole, R. V. Ecological networks and their fragility. *Nature* **442**, 259–264 (2006).
49. Long, R. A. & Azam, F. Antagonistic interactions among marine pelagic bacteria. *Appl. Environ. Microbiol.* **67**, 4975–4983 (2001).
50. Davidov, Y., Friedjung, A. & Jurkevitch, E. Structure analysis of a soil community of predatory bacteria using culture-dependent and culture-independent methods reveals a hitherto undetected diversity of *Bdellovibrio*-like organisms. *Environ. Microbiol.* **8**, 1667–1673 (2006).
51. Fuhrman, J. & Hagström, Å. in *Microbial Ecology of the Oceans* (ed. Kirchman, D.) 45–90 (Wiley, 2008).
52. Not, F., Gausling, R., Azam, F., Heidelberg, J. F. & Worden, A. Z. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ. Microbiol.* **9**, 1233–1252 (2007).
53. Massana, R. *et al.* Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl. Environ. Microbiol.* **70**, 3528–3534 (2004).
54. Countway, P. D. *et al.* Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). *Environ. Microbiol.* **9**, 1219–1232 (2007).
55. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nature Rev. Microbiol.* **5**, 801–812 (2007).
56. Angly, F. E. *et al.* The marine viromes of four oceanic regions. *PLoS Biol.* **4**, e368 (2006).
57. Bench, S. R. *et al.* Metagenomic characterization of Chesapeake bay viroplankton. *Appl. Environ. Microbiol.* **73**, 7629–7641 (2007).
58. Achtman, M. & Wagner, M. Microbial diversity and the genetic nature of microbial species. *Nature Rev. Microbiol.* **6**, 431–440 (2008).
- This paper brings together many viewpoints on the definitions and nature of microbial species.**
59. Koepf, A. *et al.* Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc. Natl Acad. Sci. USA* **105**, 2504–2509 (2008).
60. Hunt, D. E. *et al.* Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**, 1081–1085 (2008).
61. Fuhrman, J. A. *et al.* Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl Acad. Sci. USA* **103**, 13104–13109 (2006).
- This paper shows the predictability of marine microbial communities studied over a period of years, implying many features similar to familiar animal and plant systems.**
62. Wommack, K. E. & Colwell, R. R. Viroplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
63. Fuhrman, J. A. Marine viruses: biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
64. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
65. Kettler, G. C. *et al.* Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* **3**, 2515–2528 (2007).
66. Thingstad, T. F. & Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**, 19–27 (1997).
67. Fuhrman, J. A. & Suttle, C. A. Viruses in marine planktonic systems. *Oceanography* **6**, 51–63 (1993).
68. Johnson, Z. I. *et al.* Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**, 1737–1740 (2006).
69. Beja, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
70. Fuhrman, J. A., Schwalbach, M. S. & Stingl, U. Proteorhodopsins: an array of physiological roles? *Nature Rev. Microbiol.* **6**, 488–494 (2008).
71. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
72. 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).
73. Quince, C., Curtis, T. P. & Sloan, W. T. The rational exploration of microbial diversity. *ISME J.* **2**, 997–1006 (2008).
74. Rusch, D. B. *et al.* The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
75. Hewson, I., 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).
76. Franklin, M. P. *et al.* Bacterial diversity in the bacterioneuston (sea surface microlayer): the bacterioneuston through the looking glass. *Environ. Microbiol.* **7**, 723–736 (2005).
77. Fuhrman, J. A. in *Manual of Environmental Microbiology* 3rd edn (eds Hurst, C. J. *et al.*) Ch. 35 (ASM Press, 2007).
78. Ramette, A. Multivariate analyses in microbial ecology. *FEMS Microbiol. Ecol.* **62**, 142–160 (2007).
79. Chao, A., Chazdon, R. L., Colwell, R. K. & Shen, T. J. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* **62**, 361–371 (2006).

Acknowledgements I thank Å. Hagström, F. Sun, J. Steele, I. Hewson, S. Naeem, J. Green and J. Brown for helping develop ideas presented here, and C. Chow, C. Horner-Devine and J. Cram for comments. Work in my laboratory was supported by the US National Science Foundation Microbial Observatories Program and grants 0527034, 0623575, 0648581, and 0703159, and by the University of Southern California Wrigley Institute for Environmental Studies.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to the author (fuhrman@usc.edu).