PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2006) **361**, 2009–2021
doi:10.1098/rstb.2006.1928
Published online 6 October 2006

Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes

Martin F. Polz[1,*], Dana E. Hunt[1], Sarah P. Preheim[1] and Daniel M. Weinreich[2]

[1]*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[2]*Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA*

Microbes in the ocean dominate biogeochemical processes and are far more diverse than anticipated. Thus, in order to understand the ocean system, we need to delineate microbial populations with predictable ecological functions. Recent observations suggest that ocean communities comprise diverse groups of bacteria organized into genotypic (and phenotypic) clusters of closely related organisms. Although such patterns are similar to metazoan communities, the underlying mechanisms for microbial communities may differ substantially. Indeed, the potential among ocean microbes for vast population sizes, extensive migration and both homologous and illegitimate genetic recombinations, which are uncoupled from reproduction, challenges classical population models primarily developed for sexually reproducing animals. We examine possible mechanisms leading to the formation of genotypic clusters and consider alternative population genetic models for differentiation at individual loci as well as gene content at the level of whole genomes. We further suggest that ocean bacteria follow at least two different adaptive strategies, which constrain rates and bounds of evolutionary processes: the 'opportuni-troph', exploiting spatially and temporally variable resources; and the passive oligotroph, efficiently using low nutrient concentrations. These ecological lifestyle differences may represent a fundamental divide with major consequences for growth and predation rates, genome evolution and population diversity, as emergent properties driving the division of labour within microbial communities.

Keywords: natural taxa; speciation; horizontal gene flow; gene pool boundaries; genome evolution

1. INTRODUCTION

> Why, if species have descended from other species by fine gradations, do we not everywhere see innumerable transitional forms? Why is not all nature in confusion, instead of the species being, as we see them, well defined?
>
> Darwin (1859)

The extent to which prokaryotic and eukaryotic microbes dominate ocean ecosystem functions is a surprisingly recent insight. Ocean microbes form tightly integrated food webs and are responsible for the lion's share of primary production and nutrient cycling. Indeed, bacteria, the smallest and most diverse of organisms, represent the major biomass component in many oceanic regions (Whitman *et al.* 1998). Their genomes encode functions which have evolved in response to biotic and abiotic environmental constraints, and reflect the intricacies of biogeochemical cycles (DeLong & Karl 2005). Over the last 20 years, it has been established that the ocean, like most natural environments, harbours enormous genetic and genomic diversity (Giovannoni & Stingl 2005). Yet, one of the central challenges that remains is the search for structure–function relationships at the level of alleles at individual genetic loci, genes within genomes, individuals within species and species within communities. Essential questions are to what extent microbial genomes are organized into functionally cohesive and evolutionarily defined populations, what are functional units beyond the single cell and how do such units originate and self-organize under different environmental constraints?

The ocean, owing to its enormous expanse, may seem an improbable environment to examine and establish structure–function relationships in microbial communities. However, several factors make planktonic microbes a better model system than those from other globally important ecosystems. First, microbial diversity has been the focus of intense studies, especially by modern, culture-independent techniques. As a result, many of the major prokaryotic groups have been identified. Although molecular techniques have shown that microbes in the ocean, like in all other major environments, are much more diverse than previously anticipated (Giovannoni & Rappé 2000), they are probably orders of magnitude less diverse than sediment and soil communities (Gans *et al.* 2005). Second, many relevant biogeochemical gradients vary over relatively

* Author for correspondence (mpolz@mit.edu).

One contribution of 15 to a Discussion Meeting Issue 'Species and speciation in micro-organisms'.

This journal is © 2006 The Royal Society

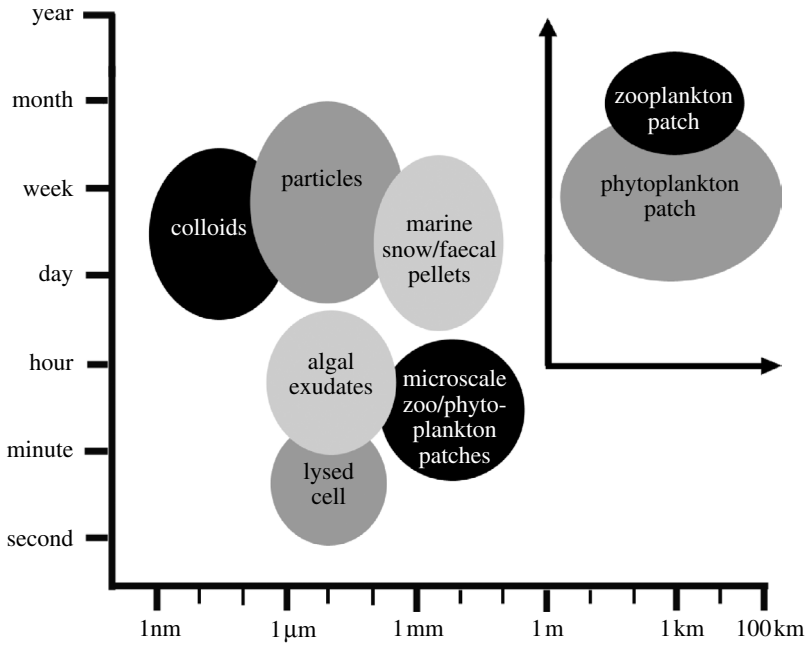

Figure 1. Estimated temporal and spatial relationships of micro- and mesoscale features in the environment affecting the growth and productivity of marine bacteria. The region to the right and above the arrows indicates features that are captured by standard oceanographic sampling methods (modified from Dickey (1991) and Seymour (2005)).

large spatial scales (metres to kilometres), but even small-scale heterogeneities will ultimately be easier to resolve in the context of aquatic environments compared to soils and sediments. Third, metagenomics is starting to be systematically applied to ocean environments and will allow correlation of major differences in genome features with environmental parameters and biogeochemical gradients (Venter *et al.* 2004; DeLong *et al.* 2006). These factors combined with the availability of several well-studied model organisms representing diverse ecological strategies hold promise for major advances in deciphering the patterns and determinants of microbial diversity.

Ocean water represents a complex and dynamic landscape of physio-chemical parameters (Goldman 1984). While it is possible to define 'average' conditions (e.g. nutrient concentration, temperature, light penetration) for large regions of the ocean, many relevant parameters show variation on much smaller scales (Azam & Ammerman 1984). Thus, when considering how bacterioplankton adapt to ecological conditions, it is important to evaluate processes at the relevant scale (figure 1). Rather than a uniform environment, the ocean is better represented as an evolving mosaic of microenvironments with varying spatial and temporal scales. While there has been some success in correlating genetic and genomic diversity with large-scale gradients, it remains largely unknown to what extent the heterogeneity on smaller scales select for genomic differentiation, which may ultimately result in population structure. We therefore begin our discussion by drawing a picture of relevant ecological parameters in the surface ocean with special emphasis on spatial and temporal heterogeneities on bacterial scales. Second, we provide an overview of the status of microbial community analysis. We emphasize recent observations of nucleotide sequence clusters and the emergence of hypotheses that these clusters represent ecologically differentiated populations. This will be followed by a theory- and observation-based critical assessment of the evolutionary origins of sequence clusters. Finally, we give a genomic perspective on population differentiation and end with considerations of the evolutionary consequences of adaptation to prevalent environmental parameters in the ocean. Overall, we focus on several well-researched bacterial groups in the photic zone, i.e. the first 100 m or so penetrated by light, where enough data on diversity and ecological constraints are available to speculate about their implications for population structure and dynamics. For more comprehensive overviews of microbial life and diversity in the ocean, we refer the reader to excellent recent reviews (Giovannoni & Rappé 2000; DeLong & Karl 2005; Giovannoni & Stingl 2005).

2. THE LIFE AQUATIC (ECOLOGICAL PARAMETERS IN THE OCEAN SURFACE)

Despite extensive and complex variation in physio-chemical parameters and large differences in nutrient status of different ocean regions, the average total cell concentration in seawater is remarkably constant. In open ocean and coastal regions, it is typically approximately 10^5 and 10^6 cells ml^{-1}, respectively (Whitman *et al.* 1998). Predation is thought to control this average since bacteria and their predators (viruses and protozoa) possess comparable reproductive rates. Prokaryotic cell numbers in the water column are thus considered to be in steady state, with the exception of bloom situations when populations can temporarily escape predation control. Thus, an important consequence of such tight control of total cell numbers is that increase in one population needs to be coupled to decrease in at least some others.

A second consequence is that communities are roughly in a steady state and an increased nutrient supply results primarily in higher turnover rather than

increases in total cell numbers. On the other hand, microbial growth rates in the ocean are relatively slow owing to the typically low nutrient concentration in bulk seawater. This is particularly true for the vast expanse of the subtropical gyres where major resources, such as nitrogen and phosphorus, are frequently too low to be measured. Although other ocean provinces, in particular coastal environments and estuaries, contain considerably higher and more variable nutrient concentrations, the ocean can overall be considered a desert-like environment. In bulk seawater, bacterial growth dynamics are improbable to ever exceed first-order kinetics. This is reflected in low doublings per day of 0.05–0.3 for open ocean environments but up to 1–2 for coastal zones (Ducklow 2000; Crump *et al.* 2004). However, ocean water is not homogeneous and locally confined higher growth rates can be achieved by bacteria clustered around or attached to small particles (Worden *et al.* 2006; figure 1). Yet such relationships are rarely captured since present analysis techniques still require sampling of relatively large volumes, which average conditions and overlook spatially small nutrient inhomogeneities.

The distinction between nutrient-poor interstitial water and nutrient-rich particles and patches probably represents a key parameter in structuring oceanic microbial communities. Bacteria inhabiting the interstitial water experience low but relatively steady nutrient concentrations, while those exploiting nutrient patches live in a temporally and spatially highly variable landscape requiring frequent abrupt changes in metabolism. This basic distinction is probably a general feature of ocean environments even though bulk nutrient concentration and particle concentration and quality can vary considerably.

The causes of small-scale nutrient microenvironments are varied and include patches of dissolved and particulate organic matter: high- and low-molecular weight algal exudates; cellular material released by viral lysis or autolysis; particles produced by sloppy zooplankton feeding; faecal pellets; marine snow; abandoned food webs; detritus; transparent exopolymers; and colloids (figure 1). Within nutrient patches, biologically labile compounds can be two to three orders of magnitude more abundant than in the surrounding water (Fenchel 2002). Although the relative importance of different patches can vary, in the surface ocean, phytoplankton are the major agents of dissolved organic matter (DOM) transfer to the heterotrophic bacterial component of food webs. Algae can reach densities of the order of 10^3 cells ml^{-1} and have been reported to exude from less than 1 to 50% of their daily photosynthate resulting in a diffusion zone enriched in DOM (Hellebust 1974; Bertilsson & Jones 2003). Similarly, organic particles (e.g. faecal pellets, marine snow) may leak DOM because the rate of polymer hydrolysis by attached bacteria exceeds monomer uptake rates. It has thus been suggested that sinking particles leave behind a wake of enriched nutrients, which can be exploited by heterotrophic bacteria. In fact, use of such patchy nutrient sources may increase the rate of mineralization in the ocean by a factor of 2, relative to a uniform nutrient scenario (Fenchel 2002), and experiments show that bacterial productivity decreases by 12–20% when seawater samples are homogenized (Moeseneder & Herndl 1995).

To what extent bacteria can actively exploit nutrient patches is strongly dependent on the spatial and temporal distributions of their nutrient sources, as well as their ability to track and cluster around them. Additionally, quality and duration of nutrient supply from individual patches is highly variable (figure 1). For example, algal cells might exude carbon monomers and polymers when sufficient light allows for photosynthesis (Bertilsson & Jones 2003); faecal pellets most probably leak more complex DOM and have been suggested to lose the majority of their DOM during the first 6 h after they are released (Urban-Rich 1999), while nutrient patches from lysed cells contain complex organic matter and may dissipate within minutes (Blackburn *et al.* 1998).

Motility and chemotaxis (i.e. the ability to sense and respond to gradients of a chemical compound) thus become valuable assets in the exploitation of these transient and localized nutrient sources. However, motility also comes at considerable metabolic cost so that motility may increase evolutionary fitness only if there is a minimum density of patches in the water column (Kiorboe *et al.* 2002; Mitchell 2002). Recent metagenomic observations indirectly support this notion; genomes from surface water have higher incidence of flagellar genes than those from deep waters where patches are at much lower concentration and consist of more recalcitrant material, so that the cost of getting from one patch to another may outweigh the benefits (DeLong *et al.* 2006). Theoretical considerations similarly support this hypothesis that cost optimization limits the adaptive value of chemotaxis (Kussel & Leiber 2005).

Given these considerations, particles and nutrient inhomogeneities probably represent a highly stochastic ecological landscape, with major consequences on bacterial distributions (figure 1). It has been observed that prokaryotic cell clusters form and dissipate within minutes (Blackburn *et al.* 1998) and that order of magnitude variability in prokaryote numbers over small sample scales exists in environmental samples (Duarte & Vaque 1992; Muller-Niklas *et al.* 1996; Seymour *et al.* 2000). Community profiling by molecular techniques also showed genetic differences in 1 µl but not in 25 µl seawater samples, indicating inhomogeneity at smaller scales (Kirchman 2001; Long & Azam 2001). Cells can also actively attach to particles to hydrolyse polymeric substances, and there has been considerable debate as to whether free-living and particle-attached bacteria constitute independent populations: some studies have found no significant difference between these groups (Martinez *et al.* 1996; Hollibaugh *et al.* 2000; Riemann & Winding 2001; Worm *et al.* 2001), while other investigators observed differences (DeLong *et al.* 1993; Acinas *et al.* 1997, 1999; Crump *et al.* 1999; Fandino *et al.* 2001; Knoll *et al.* 2001; Moeseneder *et al.* 2001).

As we will argue below, adaptive strategies tuned either to exploitation of the low-nutrient bulk water or high-nutrient patches may represent a fundamental divide among ecological strategies, with major consequences for growth and predation rates, genome

evolution and, ultimately, population diversity and structure. First, we review relevant knowledge of bacterioplankton diversity and then consider what may represent an ecologically or evolutionarily coherent bacterial population.

3. PATTERNS OF MICROBIAL DIVERSITY

Studies of microbial diversity have made considerable progress in recent years and are increasingly informed by population genetics and comparative genomics. What presently emerge are testable hypotheses of how to recognize ecologically differentiated populations. This is possible through observation of fine-scale patterns in bacterial community structure coupled to formulation of hypotheses regarding their origins, which will ultimately motivate establishment of environmental correlates on appropriate spatio-temporal scales.

Microbial diversity in the ocean, like most environments, has primarily been studied by assessing sequence diversity of 16S rRNA genes ('ribotypes') retrieved by PCR amplification, cloning and sequencing as a proxy for organismal diversity (Rappé & Giovannoni 2003). This has yielded at least 52 phylogenetically broadly defined bacterioplankton phyla, of which half have no cultivated representatives (Giovannoni & Stingl 2005). Novel culturing techniques and metagenomic approaches are revealing some of the features of these previously unknown types (Rappé *et al.* 2002; Venter *et al.* 2004; DeLong *et al.* 2006); however, the vast majority of bacteria in the ocean remain inaccessible to these techniques so that diversity estimation will rely on clone libraries for some time to come.

One basic observation (and problem) has been that microbial communities contain so much sequence diversity that clone libraries contain almost exclusively unique sequences. Some of this is certainly owing to PCR-induced sequencing artefacts. For example, we have shown that the fraction of unique sequences was reduced from 76 to 48% when sequence artefacts were constrained in a large 16S rRNA library (Acinas *et al.* 2004, 2005). However, this is still a high fraction of unique sequences considering that the 16S rRNAs are evolutionarily highly conserved. Similarly (PCR-amplification independent), shotgun sequencing of Sargasso Sea bacterioplankton detected 643 unique sequence types among 1412 rRNA genes using a 99% similarity cut-off to define unique sequence types (Venter *et al.* 2004).

Despite such low redundancy, important patterns of distribution of different phylogenetic groups have emerged from clone library sequence analysis and some fingerprinting methods. It is customary to lump sequences into phylotypes according to varying sequence cut-offs and then to compare the distribution of phylotypes among different samples. This has shown that many such phylotypes occur in both open ocean and coastal environments, but apparently not every phylotype is found everywhere (at least not in the same proportions; Giovannoni & Rappé 2000). For example, it was from clone libraries that the first members of the SAR11 group were identified; these, now named *Pelagibacter*, comprised at least 16% of total cells in the mesopelagic and up to 50% of bacterioplankton in the surface ocean (Morris *et al.* 2002; Malmstrom *et al.* 2005). Moreover, it has generally been found that easily cultivable phylotypes are frequently not the most abundant members of bacterioplankton, although abundance may in some cases underestimate importance owing to higher turnover rates (Worden *et al.* 2006). The notable exception is the *Roseobacter* clade, of which some groups are easily cultivable and can account for upwards of 20% of coastal bacterioplankton cells (Buchan *et al.* 2005).

Although many phylotypes show differential distribution in environmental samples, a central question has been how sequences should be grouped to allow identification of ecologically distinct populations. Some metabolic guilds of bacteria carry rRNA signatures (e.g. sulphate-reducing bacteria, methanotrophs, nitrifiers), but most metabolic or physiological functions have yet to be linked to clearly delineated phylogenetic groups (Pernthaler & Amann 2005). Thus to identify cohesive ecological populations in clone libraries, two basic approaches appear possible. On the one hand, *a priori* (or arbitrary) sequence cut-offs might be defined, corresponding to the thresholds that have been seen to correlate to known taxonomic units in other clades. On the other hand, it may be possible to search for emergent hierarchical patterns of variation among communities (i.e. naturally occurring clustering in sequence diversity), which can be interpreted based on evolutionary theory.

The first approach led to the use of 16S rRNA cut-offs at 3% sequence divergence to delineate taxonomic units; this is based on data suggesting that above 70% DNA–DNA hybridization (i.e. the traditional though theoretically dubious species cut-off), no 16S rRNA similarities of less than 97% have been found (Stackebrandt & Goebel 1994; Rosselló-Mora & Amann 2001). However, the notion of using 3% sequence cut-offs as ecologically cohesive units has been repeatedly challenged on both empirical and theoretical (see below) grounds. Most recently, analysis of such closely related strains has shown that they can have diverse and apparently ecologically differentiated genomes, suggesting that the traditional definition is far too broad. This was first discovered for pathogenicity determinant genes, which are frequently clustered in genomic islands unique to otherwise largely homogeneous pathogen genomes (Hacker & Carniel 2001). Indeed, most bacterial genomes contain a number of such differentiating islands. For example, comparative analysis of *E. coli* strains has revealed that typically hundreds of genes are unique to a given strain (Welch *et al.* 2002).

The second approach led to the search for naturally hierarchical units that requires datasets large enough to examine relationships at multiple levels of differentiation. This has only recently become possible but has already revealed fine-scale patterns of differentiation within ribotype sequences, which suggest prevalence of natural clusters with 1% internal sequence divergence. In both coastal bacterioplankton and marsh sediment sulphate-reducing bacteria samples, most sequences fell into such microdiverse sequence clusters indicating predominance of closely related

taxa (Acinas *et al.* 2004; Klepac-Ceraj *et al.* 2004). Indeed, it has been proposed that sequence clusters may represent natural units of differentiation equivalent to populations or species (Cohan 2002). But importantly, note that the numeric value of genetic diversity corresponding to observed clusters may probably vary from taxon to taxon. In §4, we evaluate present theories of how clusters may arise and thus their probable ecological significance.

4. SEQUENCE CLUSTERS AS POPULATIONS OR SPECIES?

It is generally accepted (indeed, rarely even remarked upon) that multicellular organisms are highly clustered phenotypically, i.e. the phenotypic variance within groups is far less than the variance between them, and this forms the basis of the vernacular, intuitive concept of species. Importantly, similar phenotypic clustering is observed in many other taxa, including bacteria (e.g. Goodfellow *et al.* 1997). However, criteria for ordering bacterial isolates into phenotypic clusters have frequently been biased by the goals of the researcher, so that many phenotypic groupings should perhaps be considered arbitrary. In particular, from this phenotypic point of view, pathogen classification has suffered from excessive splitting; for example, *Shigella* is now considered to be merely a variant of *E. coli* differentiated by a few traits, which have arisen independently multiple times (Pupo *et al.* 2000; Fukiya *et al.* 2004).

How are phenotypic clusters manifested genetically? Recent methods for classification of strains into populations and species have focused on the discovery of sequence clusters. In particular, multilocus sequence analysis (MLSA), which has grown out of the typing of pathogenic strains, may hold promise for the search for functionally defined populations and species (Maiden *et al.* 1998; Gevers *et al.* 2005). Since this approach targets multiple putatively neutral loci within each bacterial genome, the phylogenetic signal obtained from the concatenated gene sequence is more robust than for a single gene (Hanage *et al.* 2006*a*). Indeed, MLSA reveals sequence clusters which are congruent with some well-defined bacterial species (Godoy *et al.* 2003; Priest *et al.* 2004), and this concept is similar to the phylogenetic species concept (Taylor *et al.* 2000).

Of fundamental interest is to what extent sequence clusters denote ecologically differentiated populations and/or species. Under the classical view developed from metazoan biology, reproductive isolation arises primarily when ecological or geographical isolation defines the boundaries of species. These boundaries, in turn, ensure that mutations that give advantageous phenotypic effects in one environment are not diluted by genetic recombination with immigrants and are thought to be responsible for the relatively larger diversity between than within species. However, the elements of this classical model are not present in many biological systems, e.g. populations developing in sympatry (or parapatry) and asexual organisms. This calls into question the generality of this mechanism for understanding the basis of phenotypic clustering. Nor is this process intrinsic to the more fundamental conception of species, as populations of organisms selectively optimized to distinct ecological opportunities.

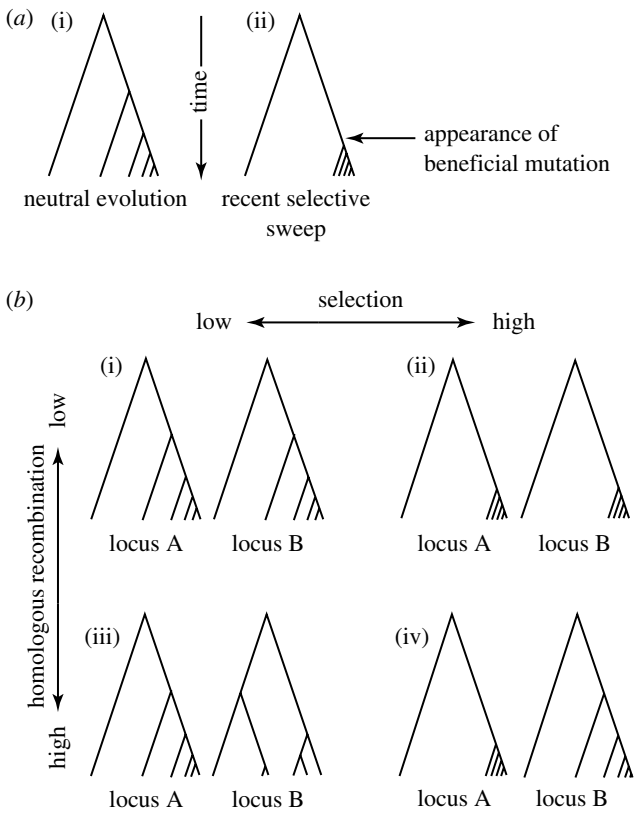

Figure 2. Schematic of the effects of selection and HR on sequence-based phylogenetic trees. (*a* (i)) In the absence of selection, branch lengths reflect the coalescent process of genetic drift. (ii) After a selective sweep, branch lengths are shortened, reflecting the loss of genetic diversity. (*b* (i, ii)) Low rates of HR between loci result in shared genealogical histories at these loci, reflected by high correlations among phylogenies. (iii, iv) Recombination disrupts this correlation, and even after a selective event, shortened branch lengths are only observed at or genetically near the target of selection.

Bacteria offer the opportunity to ask the more fundamental question: can clusters arise as a consequence of ecological specialization? One possible mechanism is that selective sweeps may periodically purge genetic variation from coexisting genomes (Cohan 2002, 2006). This assumes that clonally reproducing bacteria will accumulate mutations, which, in rare cases, are adaptive. The carrier of such adaptive mutations will increase in frequency until it has outcompeted all other strains within its niche. In strictly clonally reproducing organisms, genetic variation would be reset to zero at all loci, since only the winning clone remains. Subsequent to the sweep, all loci will begin to diversify and similar patterns of clustering should be apparent at most housekeeping loci (figure 2*b*(ii)). Importantly, variation within a niche-specific cluster would then persist because competition is not strong enough to purge variants from within the cluster. Such clusters have been termed ecotypes (Cohan 2002). One of the attractive features of the ecotype concept is that it would indeed give ecological meaning to sequence clusters observed in environmental clone libraries (Acinas *et al.* 2004).

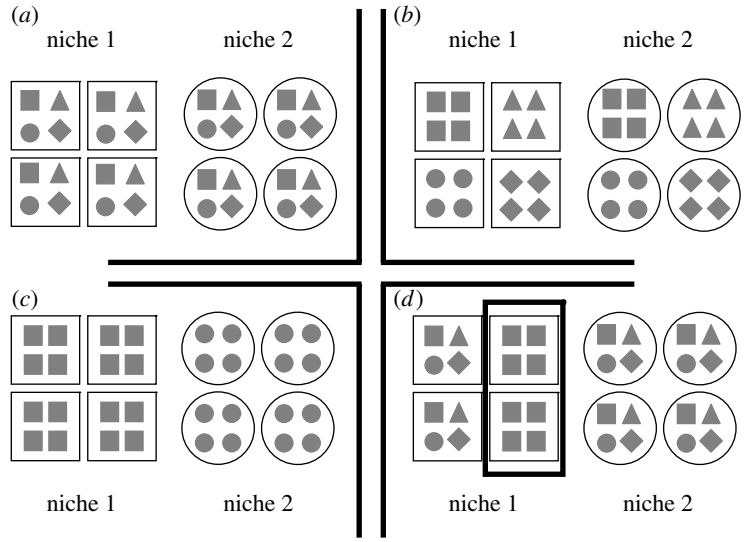

Figure 3. Idealized environmental distribution of sequence clusters assuming different degrees of ecological differentiation and/or stochastic processes of niche colonization. (*a*) Random distribution across niches with no apparent fitness differences among genotypes. (*b*) Clonal expansion within local niches owing to population bottlenecks or founder effects leading to apparent population structure. (*c*) Strong correlation with niche space indicating fitness differences. (*d*) Special case: microepidemics create a clonal expansion such that one genotype dominates in a localized area (bold line). Large boxes and circles denote distinct niche spaces; different symbols represent distinct populations within each niche space whereas the same symbol denotes individual strains from within the same sequence cluster.

The main critique for the potential of sweeps leading to clustering that can be observed using any locus arises from considerations of gene transfer among bacterial lineages (figure 2*b*(iv)). Bacteria reproduce clonally and gene transfer is an episodic event, which typically affects only small genome regions and is decoupled from reproduction. All genetic exchange in bacteria is therefore horizontal and takes the form of homologous or illegitimate recombination, where new alleles or loci arrive in the cell via transformation, transduction or conjugation. As we will argue later, each of these processes follows different rules, at least some of which will be ecologically determined so that the expected variation among bacterial genomes may be contingent on their lifestyle in the wild.

Cluster formation may be governed by a complex interplay of homologous recombination (HR) and selection, and may result in distinct patterns for different loci and genome regions (figure 2*b*(iv)). For example, HR can blur boundaries of nascent clusters if gene transfer rates are high and fitness differences between competitors low (slowing the rate of genetic homogenization in the population). In this case, both selectively favoured and neutral alleles can move among genomes, thereby diversifying clusters. In the other extreme, high selection and low recombination rates lead to entirely clonal populations after a selective sweep (figure 2*b*(ii)). However, neither HR rates nor fitness differences are presently well characterized among natural microbial populations. In particular, even relatively large fitness differences for resource acquisition could be depressed by negative frequency-dependent selection. For example, phage predation can disproportionately remove the winner of competitive events (Thingstad 2000), offsetting apparent fitness advantages. Moreover, modelling has recently suggested that clusters may arise at least transiently in the absence of positive selection assuming high rates of HR (Falush *et al.* 2006; Hanage *et al.* 2006*b*); if these early results are confirmed, then some observed clusters may bear little to no ecological information.

On the other hand, once clusters are formed, they may indeed become strongly genetically isolated from each other. It has been shown in pure culture experiments that the probability of HR decreases log-linearly with sequence divergence (Roberts & Cohan 1993; Vulic *et al.* 1997; Majewski *et al.* 2000). This means that HR between divergent clusters may be so improbable that they are effectively sexually isolated (Dykhuizen & Green 1991). Patterns from MLSA are indeed consistent with this expectation. The majority of strains typically fall within defined clusters but some do not, possibly as a result of occasional introduction of divergent alleles by HR (Hanage *et al.* 2005).

Theoretical considerations suggest that divergent clusters may also form gradually by horizontal gene transfer by illegitimate recombination, because insertion of novel genetic material prevents HR in adjacent genome regions (Lawrence 2002). In this way, genetic isolation may propagate neutrally through the genome via accumulation of point mutations further inhibiting HR (Vetsigian & Goldenfeld 2005). Such processes may occur at different rates in different genomic regions (Gavrilets & Vose 2005), but they may continue until clades become essentially genetically isolated throughout their genomes (Lawrence 2002). Clusters originating via this process would then be free to diverge ecologically while carrying considerable genotypic diversity.

In order to decide whether sequence clusters arise neutrally or represent ecologically differentiated populations, we must correlate them to relevant environmental parameters or niches. Figure 3 represents an idealized scenario for expected environmental distribution of different clusters assuming varying levels of adaptation to different niches. A number of studies have indeed succeeded in establishing that organisms denoted by different clusters show different distribution

and dynamics within the same bacterioplankton communities. This was first described for SAR11 (*Pelagibacter*) clades in environmental clone libraries of 16S rRNA genes, where different depth distribution for two clusters was established (Field *et al.* 1997). A third cluster, which could initially not be correlated to environmental factors, was later discovered to have distinct temporal occurrence patterns probably triggered by stratification of the water column (Morris *et al.* 2005). Similarly, in the cyanobacterium *Prochlorococcus*, sequence clusters containing high- and low-light-adapted strains show distinct depth distribution (Moore *et al.* 1998; Rocap *et al.* 2003). Additionally, six clades denoted by differences in internal transcribed spacer (ITS) sequences displayed distinct distribution patterns on ocean-scale gradients (Johnson *et al.* 2006). In particular, temperature correlated with occurrence and tolerance limits of different isolates from within the clusters, but other ecological factors also showed a relationship (Bouman *et al.* 2006; Johnson *et al.* 2006). Temperature was also identified as a key regulator in analysis of coexisting *Vibrio* populations identified as microdiverse 16S rRNA clusters. These showed distinct shifts in population abundance between cold and warm seasons in a year long study of a temperate estuary (Randa *et al.* 2004; Thompson *et al.* 2004).

Sequence clusters are dynamic entities on evolutionary time-scales and the genes used to distinguish any clusters must have an adequate level of genetic variation. Although we recently established that clusters are a general phenomenon within a coastal bacterioplankton community where clusters were on average less than 1% divergent in 16S rRNA gene sequences (Acinas *et al.* 2004), such community averages cannot be universally applied to all clades. In other words, some clusters may be more ancient and thus be visible in conserved genes like 16S rRNAs; some may have originated more recently and thus will only be apparent using more rapidly evolving genes. For example, *Roseobacter* 16S rRNA gene sequences grouped into 99% similarity clusters showed that some of these groupings were strongly related to the environment in which these sequences were obtained (e.g. polar environments, eukaryote-associated), but overall the 16S rRNA gene was not sufficiently discriminatory to allow functional-based grouping (Buchan *et al.* 2005). Similarly, in the diverse cyanobacterial group *Prochlorococcus*, clusters are apparent in 16S rRNA genes, but clusters in the ITS correspond better to ecological differentiation (Rocap *et al.* 2002; Johnson *et al.* 2006).

Moreover, all clusters are not equally informative with respect to environmental correlations. As detailed above, some clusters may arise by neutral processes while others may originate by selective sweeps. Indeed, MLSA datasets, which are based on protein-coding sequences, typically reveal hierarchies of clusters, i.e. clusters within clusters. Which clusters correspond to ecologically differentiated genomes must be decided by correlation of each cluster hierarchy with relevant environmental parameters and genomic diversity in a population genetic framework. And ultimately what will be sought are detailed mechanistic explanations linking spatial and temporal variations in physico-chemical gradients with particular genetic elements conferring selective advantages.

5. TOWARDS BACTERIAL POPULATION GENOMICS IN THE OCEAN

The two theoretical extremes outlined, that clusters arise either by frequent sweeps (figure 2) or neutral processes (e.g. Fraser *et al.* 2005), suggest different expectations of genomic diversity. In the first case, clusters should contain relatively homogeneous genomes; in the second case, variation within clusters should be unevenly distributed among genome regions and genetic isolation may arise by the suppression of HR mediated by sequence divergence alone.

Two approaches are presently being applied to decipher patterns of sequence variation among phylogenetically closely related genomes (approx. equivalent to clusters). For the highly abundant but poorly cultivable *Prochlorococcus* and *Pelagibacter*, genome sequences of individual strains have been compared with metagenomic libraries (Giovannoni *et al.* 2005; Coleman *et al.* 2006). For less abundant but more easily cultivable organisms, such as *Vibrio*, genotypic information can be mapped onto the populations by isolation of coexisting strains (Thompson *et al.* 2005). This has the advantage that information on the individual (strain) within a population is obtained but bears the obvious danger that important types are missed owing to culture bias; culture-independent verification is thus important (Thompson *et al.* 2005). The metagenomic approach does not suffer from isolation bias (although certain genes and genomic regions may be missed owing to cloning bias); however, in these analyses, linkage across loci is ambiguous so that genomes cannot be assembled. Thus, only a population average can be obtained, and while the existence of clusters can be established for individual loci (or genome regions if large-insert bacterial artificial chromosome (BAC) cloning is employed), they cannot be mapped to entire genomes.

Both metagenomics and culturing have suggested that protein-coding genes form clusters in coexisting *Prochlorococcus*, *Pelagibacter* and *Vibrio* and that these contain considerable synonymous sequence divergence (Giovannoni *et al.* 2005; Thompson *et al.* 2005; Coleman *et al.* 2006). Thus if sweeps generate these clusters, they must occur relatively rarely. The average nucleotide identity level ranges from 78 to 95% (Giovannoni *et al.* 2005; Thompson *et al.* 2005; Coleman *et al.* 2006), suggesting that the clusters are not of recent origin. As pointed out by Stingl and Giovannoni, such rarity of sweeps is in apparent contrast to the Kimura postulate that even minor fitness differences should sweep effectively through large populations, in which case bacterioplankton with their potentially enormous effective population sizes should be continually perfected by selection (Giovannoni & Stingl 2005). On the other hand, in the vast expanse of the ocean, sweeps may take considerable time so that genomes may diversify as they sweep. Therefore, resolution of these problems will require better definition of effective population sizes, which may indeed be much smaller than the immense census sizes.

A further factor, which is presently poorly understood, is genome diversification by illegitimate recombination and gene loss. In general, comparison of closely related genomes has revealed very high heterogeneity in gene content. This was first shown in three *E. coli* strains, which surprisingly shared only approximately 40% of their combined gene complements (Welch *et al.* 2002), and this has also been recently documented in bacterioplankton species (Thompson *et al.* 2005; Coleman *et al.* 2006). Such observations have led to proposed division of genomes into sets of core (shared by all within a group) and flexible (unique to some members of a group) genes (Hacker & Carniel 2001; Lan & Reeves 2001). The flexible genome represents the balance between illegitimate recombination and deletions and has been suggested to comprise up to approximately 20% of genes in genomes (Hacker & Carniel 2001). On the other hand, the core genome is thought to be a stable complement of genes, such as ribosomal and housekeeping genes. This core reflects overall evolutionary history of the lineage, since lateral gene transfer across wide phylogenetic bounds appears rare (Lan & Reeves 2001; Daubin *et al.* 2003; Acinas *et al.* 2004).

Illegitimate recombination into the flexible genome appears to be the most important source of evolutionary innovation in bacterial genomes, since it can introduce adaptive loci. This was first discovered in the context of pathogenesis (Hacker & Carniel 2001), where the flexible genome has been implicated in niche differentiation and host adaptation. In free-living bacteria, it may help to maintain a mobile gene pool that increases fitness under specific environmental conditions (Hacker & Carniel 2001; Coleman *et al.* 2006). Moreover, illegitimate recombination has been shown to be responsible for novel functions within metabolic networks (Pal *et al.* 2005).

To what extent can new genes transferred by illegitimate recombination persist in genomes if they are not adaptive? Genomes must be able to tolerate a certain amount of non-functional gene content without detriment, since even adaptive genes may not be immediately fully functional and may undergo periods of amelioration and/or acclimation. However, deletion rates for unused genes must be roughly matched to illegitimate recombination rates since genome size does not grow without bound. Indeed, modelling has suggested that horizontally acquired sequences can persist for a long time in a substantial fraction of individuals within a bacterial population even when they are neutral or slightly deleterious (Berg & Kurland 2002; Novozhilov *et al.* 2005). Consequently, a microbial population is expected to have a large diversity of transient neutral gene content (Berg & Kurland 2002). This expectation fits observations of high gene deletion rates in non-selective environments (Nilsson *et al.* 2005), and very large genome size and gene content differences among closely related bacteria (Welch *et al.* 2002; Thompson *et al.* 2005).

The need for genomes to tolerate arrival of new genes may select for their being channelled into dedicated genomic regions and for mechanisms of re-establishment of function after loss. Indeed, many bacteria maintain plasmids, have lytic and lysogenic phages, and have large genomic regions (e.g. integrons) that can capture (novel) genes (Faruque *et al.* 1999; Rowe-Magnus *et al.* 2001; Seguritan *et al.* 2003; Dunn *et al.* 2005; Purdy *et al.* 2005). On the other hand, extrachromosomal elements (like plasmids and phages) have their own evolutionary 'agenda', which may lead to higher transfer rates of specific types of genes. One such example of channelization may be photosynthesis genes transferred by phages in *Prochlorococcus*. These genes are carried by phage and have been suggested to increase their fitness during infection by increasing gene dosage for proteins with extremely high turnover in the host cell (Lindell *et al.* 2005). However, the phage may also act as a highly efficient gene transfer agent, which may spread alleles within or among populations that are adaptive from the host's point of view.

What fraction of genes in extrachromosomal elements and genomic islands contain adaptive genes remains unknown. Further, rates and bounds of such transfer processes have not been sufficiently constrained, and so their evolutionary importance with respect to niche specialization has not been addressed within a population genetic framework. As we argue in §6, recent data suggest that lifestyle may have strong feedback on genomic mode of gene transfer and genomic diversity.

6. GENOMIC CONSEQUENCES OF ADAPTATION TO ENVIRONMENTAL VARIATIONS

In addition to these general considerations, lifestyle in the wild may have significant influence on genomic diversity. Although only few examples of comparative genomics of closely related bacterioplankton groups exist, these have already yielded some striking differences.

The ocean represents a landscape of low- and high-nutrient conditions on the microscale (figure 1). Moreover, vast regions of the open ocean contain such low bulk concentrations of major nutrients that they are difficult to measure. As detailed above, exploitation of high- and low-nutrient conditions requires different adaptations. *Pelagibacter* and *Prochlorococcus* represent one extreme among bacterioplankton, as they can exploit low-nutrient conditions so effectively that they reach numerical dominance in the open ocean. They grow relatively slowly but steadily as single, non-motile cells, which are probably rarely in contact with each other. On the other hand, *Vibrio* and *Roseobacter* (and many other fast-growing bacterioplankton) are highly motile and can move among or attach to nutrient sources. Moreover, they can exploit many alternative niches and have been detected in sediments, and in and on animals; many also have pathogenic variants. In the planktonic lifestyle, these organisms probably grow in bursts, which are locally quickly erased by predation so that overall they should have higher turnover rates than other bacterioplankton (Mourino-Perez *et al.* 2003; Worden *et al.* 2006). There are strong indications that both *Vibrio* and *Roseobacter* sense and respond to their surroundings by several mechanisms, including: quorum-sensing systems (Gram *et al.* 2002; Moran *et al.* 2004); production of antibacterial compounds (Bruhn *et al.* 2005); chemotaxis (Miller *et al.* 2004;

McCarter 2006); association with animal or algal cells (Buchan *et al.* 2005); and rapid surface colonization (Dang & Lovell 2000; Thompson & Polz 2006).

Life under conditions of extreme nutrient limitation makes metabolic efficiency and energy conservation a highly adaptive trait. Indeed, both *Prochlorococcus* and *Pelagibacter* have small (approx. 2 and 1.3 Mbp, respectively) and apparently efficiently organized genomes; for example, the latter has the shortest intergenic spacer regions known. On the other hand, *Vibrio* and *Roseobacter* have relatively large genomes (approx. 4–5 Mbp). Thus, a major difference among these two ecological types may be that growth efficiency optimization triggers lower 'tolerance' towards carriage of unused or rarely used genetic material.

Aside from possible genome size optimization, there may be significant differences in gene transfer potential. Both *Prochlorococcus* and *Pelagibacter* so far appear to lack plasmids and transposons, and integrative phages seem also rare or even absent. Life as single, free-floating cells also probably eliminates transformation as an important gene transfer mechanism since total free DNA has been shown to be at low concentration in bulk seawater (0.06–0.6 ng ml^{-1}; Karl & Bailiff 1989). Thus, lytic phages may be the only effective gene transfer agent. Vibrios appear to have many more established means of gene transfer. Indeed, vibrios typically devote approximately 1% of their gene complement to recombinases/integrases, while in *Prochlorococcus* and *Pelagibacter* less than 0.1% of genes fall into these categories (S. C. Acinas & M. F. Polz 2005, unpublished observations). For example, in *Vibrio cholerae*, the role of integrative phage in pathogenesis has been well established, and almost all vibrios appear to have large integrons. These possess genomic integrases, which can capture genes and assemble regions up to 125 kb (Heidelberg *et al.* 2000; Boucher & Stockes 2006). In *V. cholerae*, it has also been recently shown that transformation can be induced by biofilm formation on chitinaceous surfaces (Meibom *et al.* 2005); this may have major consequences for both rates of homologous and illegitimate recombination among co-occurring strains in nature.

These constraints on genome optimization and gene transfer may explain differences in observed number and extent of variable genomic islands between *Prochlorococcus* and *Vibrio*. In *Vibrio*, genomic islands can be numerous (e.g. at least 14 in *V. vulnificus*; Quirke *et al.* 2006) and are associated with phages, transposons and integrons. *Roseobacter* species have large fractions of their genome encoded on plasmids (10% in *Silicibacter pomeroy*), including important metabolic genes (Moran *et al.* 2004). Moreover, strain-to-strain variation in genome size and gene content within a natural population of *Vibrio* appears to be high. We have shown that within a *Vibrio splendidus* population, defined as a cluster of less than 1% 16S rRNA divergence, genome size variants with differences of up to 20% coexist (Thompson *et al.* 2005). Indeed, the average concentration of a unique genotype defined on the basis of gene content in the bacterioplankton samples appeared so low that the presence of a unique gene must have negligible importance on individual fitness or overall population function. On the other hand, *Prochlorococcus* MIT9312 has recently been shown by comparison with metagenomic libraries to contain only five major variable island regions, which comprise 10% of the genome. These have indications of phage origin and contain genes of which at least some are differentially expressed under different types of stress; however, it remains unknown whether they confer fitness under these conditions (Coleman *et al.* 2006).

Overall, these major differences may have important consequences for evolution and adaptation in these bacterioplankton groups. *Vibrio* and *Roseobacter* are 'opportuni-trophs' with versatile lifestyles which may necessitate flexible genomes. Indeed, genes may be adaptive under one ecological circumstance but (nearly) neutral under another. For example, some genes expressed in a fish gut may remain unused when exploiting algal exudates. This may indeed explain the high genotypic diversity of vibrios encountered in bacterioplankton (Thompson *et al.* 2005). On the other hand, genome optimization in *Prochlorococcus* and *Pelagibacter* may limit their adaptability, since gene import and presence of (frequently) unused genes may have much stronger negative fitness effect. Finally, their exclusively single-cell lifestyle may further limit avenues of gene exchange and may lead to genomes being more similar within clusters.

7. CONCLUSIONS

Microbes dominate marine biomass and are key players in nutrient cycling and primary production in the ocean. Although microbial diversity has been studied extensively, there is still little theoretical understanding or experimental evidence of ecologically coherent groupings in the wild. Nonetheless, advances in microbial ecology, genomics and evolution promise to yield insights into structure–function relationships in microbial communities. An important first step will be the coalescence of theory and observation of genotypic (and phenotypic) clusters within microbial communities. In order to decide whether such clusters represent ecologically differentiated populations, their dynamics will have to be correlated with distinct environmental compartments (e.g. zooplankton, particles, the microzone around algae) and environmental factors (e.g. temperature, salinity, light) at appropriate spatio-temporal scales (figures 1 and 3). Increasing genomic and metagenomic data from closely related organisms will also allow development of mechanistic understanding of how these clusters develop by testing the theoretical models for their consistency with environmental data. Although population genetic patterns such as those illustrated in figure 2 are consistent with the action of selective sweeps, one cannot rule out other processes, such as demographics including migration and locally confined bursts (e.g. microepidemics; Fraser *et al.* 2005; figure 3). Just as selective and non-selective processes can skew gene genealogies away from the neutral coalescent expectation, so too can they skew expectations for variation in gene content. Finally, many of these fundamental challenges of developing functional mapping from genetics and genomics to ecological and evolutionary differences are not unique to microbes,

but rather represent some of the central problems in biology. We are optimistic that work in this field will successfully lead to answers that were first posed by Darwin almost 150 years ago.

This work was supported in part by grants from the DOE Genomes to Life program, NSF Biological Oceanography and the NSF/NIEHS sponsored Woods Hole Centre for Ocean and Human Health.

REFERENCES

Acinas, S. G., Rodríguez-Valera, F. & Pedrós-Alió, C. 1997 Spatial and temporal variation in marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16S rDNA. *FEMS Microbiol. Ecol.* **24**, 27–40. (doi:10.1111/j.1574-6941.1997.tb00420.x)

Acinas, S. G., Anton, J. & Rodriguez-Valera, F. 1999 Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **65**, 514–522.

Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L. & Polz, M. F. 2004 Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**, 551–554. (doi:10.1038/nature02649)

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V. & Polz, M. F. 2005 PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* **71**, 8966–8969. (doi:10.1128/AEM.71.12.8966-8969.2005)

Azam, F. & Ammerman, J. W. 1984 Cycling of organic matter by bacterioplankton in pelagic marine ecosystems: microenvironmental considerations. In *Flows of energy and materials in marine ecosystems* (ed. M. J. R. Fasham), pp. 345–360. New York, NY: Plenham Press.

Berg, O. G. & Kurland, C. G. 2002 Evolution of microbial genomes: sequence acquisition and loss. *Mol. Biol. Evol.* **19**, 2265–2276.

Bertilsson, S. & Jones, J. G. 2003 Supply of dissolved organic matter to aquatic ecosystems: autochthonous sources. In *Aquatic ecosystems: interactivity of dissolved organic matter* (ed. S. E. G. Findlay & R. L. Sinsabaugh), pp. 3–24. San Diego, CA: Academic Press.

Blackburn, N., Fenchel, T. & Mitchell, J. 1998 Microscale nutrient patches in planktonic habitats shown by chemotactic bacteria. *Science* **282**, 2254–2256. (doi:10.1126/science.282.5397.2254)

Boucher, Y. & Stockes, H. W. 2006 The roles of lateral gene transfer and vertical descent in *Vibrio* evolution. In *The biology of Vibrios* (ed. F. L. Thompson, B. Austin & J. Swings), pp. 84–94. Washington, DC: ASM Press.

Bouman, H. A. *et al*. 2006 Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**, 918–921. (doi:10.1126/science.1122692)

Bruhn, J. B., Nielsen, K. F., Hjelm, M., Hansen, M., Bresciani, J., Schulz, S. & Gram, L. 2005 Ecology, inhibitory activity, and morphogenesis of a marine antagonistic bacterium belonging to the *Roseobacter* clade. *Appl. Environ. Microbiol.* **71**, 7263–7270. (doi:10.1128/AEM.71.11.7263-7270.2005)

Buchan, A., Gonzalez, J. M. & Moran, M.-A. 2005 Overview of the marine roseobacter lineage. *Appl. Environ. Microbiol.* **71**, 5665–5677. (doi:10.1128/AEM.71.10.5665-5677.2005)

Cohan, F. M. 2002 What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487. (doi:10.1146/annurev.micro.56.012302.160634)

Cohan, F. M. 2006 Toward a conceptual and operational union of bacterial systematics, ecology and evolution. *Phil. Trans. R. Soc. B* **361**, 1985–1996. (doi:10.1098/rstb.2006.1918)

Coleman, M., Sullivan, M. B., Martiny, A. C., Steglich, C., Barry, K., DeLong, E. F. & Chisholm, S. W. 2006 Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**, 1768–1770. (doi:10.1126/science.1122050)

Crump, B. C., Armbrust, E. V. & Baross, J. A. 1999 Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean. *Appl. Environ. Microbiol.* **65**, 3192–3204.

Crump, B. C., Hopkinson, C. S., Sogin, M. L. & Hobbie, J. E. 2004 Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. *Appl. Environ. Microbiol.* **70**, 1494–1505. (doi:10.1128/AEM.70.3.1494-1505.2004)

Dang, H. Y. & Lovell, C. R. 2000 Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. *Appl. Environ. Microbiol.* **66**, 467–475. (doi:10.1128/AEM.66.2.467-475.2000)

Darwin, C. 1859 The origin of species by means of natural selection, or the preservation of favored races in the struggle for life. London: John Murray.

Daubin, V., Moran, N. A. & Ochman, H. 2003 Phylogenetics and the cohesion of bacterial genomes. *Science* **301**, 829–832. (doi:10.1126/science.1086568)

DeLong, E. F. & Karl, D. M. 2005 Genomic perspectives in microbial oceanography. *Nature* **439**, 322–325.

DeLong, E. F., Franks, D. G. & Alldredge, A. L. 1993 Phylogenetic diversity of aggregate-attached vs. free-living marine bacteria assemblages. *Limnol. Oceanogr.* **38**, 924–934.

DeLong, E. F. *et al*. 2006 Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496–503. (doi:10.1126/science.1120250)

Dickey, T. D. 1991 The emergence of congruent high-resolution physical and biooptical measurements in the upper ocean and their applications. *Rev. Geophys.* **29**, 383–413.

Duarte, C. M. & Vaque, D. 1992 Scale dependence of bacterioplankton patchiness. *Mar. Ecol. Prog. Ser.* **84**, 95–100.

Ducklow, H. W. 2000 Bacterial production and biomass in the oceans. In *Microbial ecology of the oceans* (ed. D. L. Kirchman), pp. 85–120. New York, NY: Wiley-Liss.

Dunn, A. K., Martin, M. O. & Stabb, E. V. 2005 Characterization of pES213, a small mobilizable plasmid from *Vibrio fischeri*. *Plasmid* **54**, 114–134. (doi:10.1016/j.plasmid.2005.01.003)

Dykhuizen, D. E. & Green, L. 1991 Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* **173**, 7257–7268.

Falush, D., Torpdahl, M., Didelot, X., Conrad, D. F., Wilson, D. J. & Achtman, M. 2006 Mismatch induced speciation in *Salmonella*: model and data. *Phil. Trans. R. Soc. B* **361**, 2045–2053. (doi:10.1098/rstb.2006.1925)

Fandino, L. B., Riemann, L., Steward, G. F., Long, R. A. & Azam, F. 2001 Variations in bacterial community structure during a dinoflagellate bloom analyzed by DGGE and 16S rDNA sequencing. *Aquat. Microb. Ecol.* **23**, 119–130.

Faruque, S. M., Rahman, M. M., Asadulghani Islam, K. M. N. & Mekalanos, J. J. 1999 Lysogenic conversion of environmental *Vibrio mimicus* strains by CTX Phi. *Infect. Immun.* **67**, 5723–5729.

Fenchel, T. 2002 Microbial behavior in a heterogeneous world. *Science* **296**, 1068–1071. (doi:10.1126/science.1070118)

Field, K. G., Gordon, D., Wright, T., Rappe, M., Urbach, E., Vergin, K. & Giovannoni, S. J. 1997 Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl. Environ. Microbiol.* **63**, 63–70.

Fraser, C., Hanage, W. P. & Spratt, B. G. 2005 Neutral microepidemic evolution of bacterial pathogens. *Proc. Natl Acad. Sci. USA* **102**, 1968–1973. (doi:10.1073/pnas.0406993102)

Fukiya, S., Mizoguchi, H., Tobe, T. & Mori, H. 2004 Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J. Bacteriol.* **186**, 3911–3921. (doi:10.1128/JB.186.12.3911-3921.2004)

Gans, J., Wolinsky, M. & Dunbar, J. 2005 Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**, 1387–1390. (doi:10.1126/science.1112665)

Gavrilets, S. & Vose, A. 2005 Dynamic patterns of adaptive radiation. *Proc. Natl Acad. Sci. USA* **102**, 18 040–18 045. (doi:10.1073/pnas.0506330102)

Gevers, D. *et al*. 2005 Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* **3**, 733–739. (doi:10.1038/nrmicro1236)

Giovannoni, S. & Rappé, M. 2000 Evolution, diversity, and molecular ecology of marine prokaryotes. In *Microbial ecology of the oceans* (ed. D. L. Kirchman), pp. 47–84. New York, NY: Wiley-Liss, Inc.

Giovannoni, S. J. & Stingl, U. 2005 Molecular diversity and ecology of microbial plankton. *Nature* **437**, 343–348. (doi:10.1038/nature04158)

Giovannoni, S. J. *et al*. 2005 Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245. (doi:10.1126/science.1114057)

Godoy, D., Randle, G., Simpson, A. J., Aanensen, D. M., Pitt, T. L., Kinoshita, R. & Spratt, B. G. 2003 Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholdefia mallei*. *J. Clin. Microbiol.* **41**, 4913–4913. (doi:10.1128/JCM.41.10.4913.2003)

Goldman, J. C. 1984 Oceanic nutrient cycles. In *Flows of energy and materials in marine ecosystems* (ed. M. J. R. Fasham). New York, NY: Plenham Press.

Goodfellow, M., Manfio, G. P. & Chun, J. 1997 Towards a practical species concept for cultivable bacteria. In *The units of biodiversity* (ed. F. Claridge, H. A. Dawah & M. R. Wilson). London, UK: Chapman and Hall.

Gram, L., Grossart, H.-P., Schlingloff, A. & Kiørboe, T. 2002 Possible quorum sensing in marine snow bacteria: production of acylated homoserine lactones by *Roseobacter* strains isolated from marine snow. *Appl. Environ. Microbiol.* **68**, 4111–4116. (doi:10.1128/AEM.68.8.4111-4116.2002)

Hacker, J. & Carniel, E. 2001 Ecological fitness, genomic islands and bacterial pathogenicity—a Darwinian view of the evolution of microbes. *EMBO Rep.* **2**, 376–381.

Hanage, W. P., Fraser, C. & Spratt, B. G. 2005 Fuzzy species among recombinogenic bacteria. *BMC Biol.* **3**, 6. (doi:10.1186/1741-7007-3-6)

Hanage, W. P., Fraser, C. & Spratt, B. G. 2006*a* Sequences, sequence clusters and bacterial species. *Phil. Trans. R. Soc. B* **361**, 1917–1927. (doi:10.1098/rstb.2006.1917)

Hanage, W. P., Spratt, B. G., Turner, K. M. E. & Fraser, C. 2006*b* Modelling bacterial speciation. *Phil. Trans. R. Soc. B* **361**, 2039–2044. (doi:10.1098/rstb.2006.1926)

Heidelberg, J. F. *et al*. 2000 DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483. (doi:10.1038/35020000)

Hellebust, J. A. 1974 Extracellular products. In *Algal physiology and biochemistry* (ed. W. D. P. Stewart), pp. 838–863. London, UK: Blackwell Scientific.

Hollibaugh, J. T., Wong, P. S. & Murrell, M. C. 2000 Similarity of particle-associated and free-living bacterial communities in northern San Francisco Bay, California. *Aquat. Microb. Ecol.* **21**, 103–114.

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S. & Chisholm, S. W. 2006 Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**, 1737–1740. (doi:10.1126/science.1118052)

Karl, D. M. & Bailiff, M. D. 1989 The measurement and distribution of dissolved nucleic acids in aquatic environments. *Limnol. Oceanogr.* **34**, 543–558.

Kiorboe, T., Grossart, H. P., Ploug, H. & Tang, K. 2002 Mechanisms and rates of bacterial colonization of sinking aggregates. *Appl. Environ. Microbiol.* **68**, 3996–4006. (doi:10.1128/AEM.68.8.3996-4006.2002)

Kirchman, D. L. 2001 Measuring bacterial biomass production and growth rates from leucine incorporation in natural aquatic environments. In *Methods in microbiology*, vol. 30 (ed. J. H. Paul). San Diego, CA: Academic Press.

Klepac-Ceraj, V., Bahr, M., Crump, B. C., Teske, A. P., Hobbie, J. E. & Polz, M. F. 2004 High overall diversity and dominance of microdiverse relationships in salt marsh sulfate-reducing bacteria. *Environ. Microbiol.* **6**, 686–698. (doi:10.1111/j.1462-2920.2004.00600.x)

Knoll, S., Zwisler, W. & Simon, M. 2001 Bacterial colonization of early stages of limnetic diatom microaggregates. *Aquat. Microb. Ecol.* **25**, 141–150.

Kussel, E. & Leiber, S. 2005 Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078. (doi:10.1126/science.1114383)

Lan, R. T. & Reeves, P. R. 2001 When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends Microbiol.* **9**, 419–424. (doi:10.1016/S0966-842X(01)02133-3)

Lawrence, J. G. 2002 Gene transfer in bacteria: speciation without species? *Theor. Popul. Biol.* **61**, 449–460. (doi:10.1006/tpbi.2002.1587)

Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. 2005 Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89. (doi:10.1038/nature04111)

Long, R. A. & Azam, F. 2001 Microscale patchiness of bacterioplankton assemblage richness in seawater. *Aquat. Microb. Ecol.* **26**, 103–113.

Maiden, C. J. *et al*. 1998 Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl Acad. Sci. USA* **95**, 3140–3145. (doi:10.1073/pnas.95.6.3140)

Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F. M. & Dowson, C. G. 2000 Barriers to genetic exchange between species: *Streptococcus pneumoniae* transformation. *J. Bacteriol.* **182**, 1016–1023. (doi:10.1128/JB.182.4.1016-1023.2000)

Malmstrom, R. R., Cottrell, M. T., Elifantz, H. & Kirchman, D. L. 2005 Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Appl. Environ. Microbiol.* **71**, 2979–2986. (doi:10.1128/AEM.71.6.2979-2986.2005)

Martinez, J., Smith, D. C., Steward, G. F. & Azam, F. 1996 Variability in ectohydrolytic enzyme activities of pelagic marine bacteria and its significance for substrate processing in the sea. *Aquat. Microb. Ecol.* **10**, 223–230.

McCarter, L. L. (ed.) 2006 *Motility and chemotaxis. The biology of vibrios*. Washington, DC: ASM Press.

Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C.-Y. & Schoolnik, G. K. 2005 Chitin induces natural competence in *Vibrio cholerae*. *Science* **310**, 1824–1827. (doi:10.1126/science.1120096)

Miller, T. R., Hnilicka, K., Dziedzic, A., Desplats, P. & Belas, R. 2004 Chemotaxis of *Silicibacter* sp strain TM1040 toward dinoflagellate products. *Appl. Environ. Microbiol.* **70**, 4692–4701. (doi:10.1128/AEM.70.8.4692-4701.2004)

Mitchell, J. G. 2002 The energetics and scaling of search strategies in bacteria. *Am. Nat.* **160**, 727–740. (doi:10.1086/343874)

Moeseneder, M. M. & Herndl, G. J. 1995 Influence of turbulence on bacterial production in the sea. *Limnol. Oceanogr.* **40**, 1466–1473.

Moeseneder, M. M., Winter, C. & Herndl, G. J. 2001 Horizontal and vertical complexity of attached and free-living bacteria of the Eastern Mediterranean Sea determined by 16S rDNA and 16S rRNA fingerprints. *Limnol. Oceanogr.* **46**, 95–107.

Moore, L. R., Rocap, G. & Chisholm, S. W. 1998 Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**, 464–467. (doi:10.1038/30965)

Moran, M. A. *et al*. 2004 Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**, 910–913. (doi:10.1038/nature03170)

Morris, R. M., Connon, S. A., Vergin, K. & Giovannoni, S. J. 2002 SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806–810. (doi:10.1038/nature01240)

Morris, R. M., Cho, J. C., Rappe, M., Vergin, K. L. & Carlson, C. A. 2005 Bacterioplankton responses to deep seasonal mixing in the Sargasso Sea. *Limnol. Oceanogr.* **50**, 382–391.

Mourino-Perez, R. R., Worden, A. Z. & Azam, F. 2003 Growth of *Vibrio cholerae* O1 in red tide waters off California. *Appl. Environ. Microbiol.* **69**, 6923–6931. (doi:10.1128/AEM.69.11.6923-6931.2003)

Muller-Niklas, G., Agis, M. & Herndl, G. J. 1996 Microscale distribution of bacterioplankton in relation to phytoplankton: results from 100-nl samples. *Limnol. Oceanogr.* **41**, 1577–1582.

Nilsson, A. I., Koskiniemi, S., Eriksson, S., Kugelberg, E., Hinton, J. C. D. & Andersson, D. I. 2005 Bacterial genome size reduction by experimental evolution. *Proc. Natl Acad. Sci. USA* **102**, 12 112–12 116. (doi:10.1073/pnas.0503654102)

Novozhilov, A. S., Karev, G. P. & Koonin, E. V. 2005 Mathematical modeling of evolution of horizontally transferred genes. *Mol. Biol. Evol.* **22**, 1721–1732. (doi:10.1093/molbev/msi167)

Pal, C., Papp, B. & Lercher, M. J. 2005 Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* **37**, 1372–1375.

Pernthaler, J. & Amann, R. 2005 Fate of heterotrophic microbes in pelagic habitats: focus on populations. *Micrbiol. Mol. Biol. Rev.* **69**, 440–461. (doi:10.1128/MMBR.69.3.440-461.2005)

Priest, F. G., Barker, M., Baillie, L. W. J., Holmes, E. C. & Maiden, M. C. J. 2004 Population structure and evolution of the *Bacillus cereus* group. *J. Bacteriol.* **186**, 7959–7970. (doi:10.1128/JB.186.23.7959-7970.2004)

Pupo, G. M., Lan, R. & Reeves, P. R. 2000 Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc. Natl Acad. Sci. USA* **97**, 10 567–10 572. (doi:10.1073/pnas.180094797)

Purdy, A., Rohwer, F., Edwards, R., Azam, F. & Bartlett, D. H. 2005 A glimpse into the expanded genome content of *Vibrio cholerae* through identification of genes present in environmental strains. *J. Bacteriol.* **187**, 2992–3001. (doi:10.1128/JB.187.9.2992-3001.2005)

Quirke, A. M., Reen, F. J., Claesson, M. J. & Boyd, E. F. 2006 Genomic island identification in *Vibrio vulnificus* reveals significant genome plasticity in this human pathogen. *Bioinformatics* **22**, 905–910. (doi:10.1093/bioinformatics/btl015)

Randa, M. A., Polz, M. F. & Lim, E. 2004 Effects of temperature and salinity of *Vibrio vulnificus* population dynamics as assessed by quantitative PCR. *Appl. Environ. Microbiol.* **70**, 5469–5476. (doi:10.1128/AEM.70.9.5469-5476.2004)

Rappé, M. S. & Giovannoni, S. J. 2003 The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**, 369–394. (doi:10.1146/annurev.micro.57.030502.090759)

Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. 2002 Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633.

Riemann, L. & Winding, A. 2001 Community dynamics of free-living and particle-associated bacterial assemblages during a freshwater phytoplankton bloom. *Microb. Ecol.* **42**, 274–285. (doi:10.1007/s00248-001-0018-8)

Roberts, M. S. & Cohan, F. M. 1993 The effect of DNA sequence divergence on sexual isolation in *Bacillus*. *Genetics* **134**, 401–408.

Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. 2002 Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* **68**, 1180–1191. (doi:10.1128/AEM.68.3.1180-1191.2002)

Rocap, G. *et al*. 2003 Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042–1047. (doi:10.1038/nature01947)

Rosselló-Mora, R. & Amann, R. 2001 The species concept for prokaryotes. *FEMS Microbiol. Rev.* **25**, 39–67. (doi:10.1016/S0168-6445(00)00040-1)

Rowe-Magnus, D. A., Guerout, A.-M., Ploncard, P., Dychinco, B., Davies, J. & Mazel, D. 2001 The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc. Natl Acad. Sci. USA* **98**, 652–657. (doi:10.1073/pnas.98.2.652)

Seguritan, V., Feng, I. W., Rohwer, F., Swift, M. & Segall, A. M. 2003 Genome sequences of two closely related *Vibrio parahaemolyticus* phages, VP16T and VP16C. *J. Bacteriol.* **185**, 6434–6447. (doi:10.1128/JB.185.21.6434-6447.2003)

Seymour, J. R. 2005 Miscoscale spatiotempioral heterogeneity of marine microbial communities, p. 244. Ph.D. thesis, Flinders University, Adelaide, South Australia, School of Biological Sciences.

Seymour, J. R., Mitchell, J. G., Pearson, L. & Waters, R. L. 2000 Heterogeneity in bacterioplankton abundance from 4.5 millimetre resolution sampling. *Aquat. Microb. Ecol.* **22**, 143–153.

Stackebrandt, E. & Goebel, B. M. 1994 Taxonomic note: a place for DNA–DNA reassociation kinetics and sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44**, 846–849.

Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S. & Fisher, M. C. 2000 Phylogenetic species recognition and species concept in fungi. *Fungal Genet. Biol.* **31**, 21–32. (doi:10.1006/fgbi.2000.1228)

Thingstad, T. F. 2000 Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **45**, 1320–1328.

Thompson, J. R. & Polz, M. F. 2006 Dynamics of *Vibrio* populations and their role in environmental nutrient cycling. In *The biology of Vibrios* (ed. F. L. Thompson, B. Austin & J. Swings), pp. 190–203. Washington, DC: ASM Press.

Thompson, J. R., Randa, M. A., Marcelino, L. A., Tomita, A., Lim, E. L. & Polz, M. F. 2004 Diversity and dynamics of a North Atlantic coastal *Vibrio* community. *Appl. Environ. Microbiol.* **70**, 4103–4110. (doi:10.1128/AEM.70.7.4103-4110.2004)

Thompson, J. R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D. E., Benoit, J., Sarma-Rupavtarm, R., Distel, D. L. & Polz, M. F. 2005 Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**, 1311–1313. (doi:10.1126/science.1106028)

Urban-Rich, J. 1999 Release of dissolved organic carbon from copepod fecal pellets in the Greenland Sean. *J. Exp. Mar. Biol. Ecol.* **232**, 107–124. (doi:10.1016/S0022-0981(98)00104-X)

Venter, C. J. *et al*. 2004 Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74. (doi:10.1126/science.1093857)

Vetsigian, K. & Goldenfeld, N. 2005 Global divergence of microbial genome sequences mediated by propagating fronts. *Proc. Natl Acad. Sci. USA* **102**, 7332–7337. (doi:10.1073/pnas.0502757102)

Vulic, M., Dionisio, F., Taddei, F. & Radman, M. 1997 Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc. Natl Acad. Sci. USA* **94**, 9763–9767. (doi:10.1073/pnas.94.18.9763)

Welch, R. A. *et al*. 2002 Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 17 020–17 024. (doi:10.1073/pnas.252529799)

Whitman, W. B., Coleman, D. C. & Wiebe, W. J. 1998 Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583. (doi:10.1073/pnas.95.12.6578)

Worden, A. Z., Seidel, M., Smriga, S., Wick, A., Malfatti, F., Bartlett, D. & Azam, F. 2006 Trophic regulation of *V. cholerae* in coastal marine waters. *Environ. Microbiol.* **8**, 21–29. (doi:10.1111/j.1462-2920.2005.00863.x)

Worm, J., Gustavson, K., Garde, K., Borch, N. H. & Sondergaard, M. 2001 Functional similarity of attached and free-living bacteria during freshwater phytoplankton blooms. *Aquat. Microb. Ecol.* **25**, 103–111.