



Molecular diversity of bacterioplankton: link to a predictive biogeochemistry of pelagic ecosystems

Manfred G. Höfle^{1,*}, David L. Kirchman², Richard Christen³, Ingrid Brettar¹

¹Department Vaccinology, Helmholtz Centre of Infection Research (HZI), Inhoffenstr. 7, 38124 Braunschweig, Germany

²College of Marine and Earth Studies, University of Delaware, Lewes, Delaware 19958, USA

³Virtual Biology Lab, Centre for Biochemistry, University of Nice Sophia Antipolis & CNRS, Campus Valrose, 06108 Nice, France

ABSTRACT: In this paper, we focus on the current understanding of the microbial diversity of bacterioplankton and its link to the biogeochemistry of the pelagic zone of aquatic ecosystems. During the last 2 decades, a large array of molecular approaches has been developed and applied to assess the diversity of bacterioplankton. One key result of the classical molecular approaches, such as community fingerprints and environmental clone libraries, is that the bacterioplankton community contains only a few abundant species and a large number of less abundant to rare members. This is in contrast to soil bacterial communities, where no abundant species are detected and the total number of species is substantially larger. New approaches, such as metagenomics and pyrosequencing, are discussed in terms of their ability to determine the diversity and abundance of the rare members of the community and the biogeochemical potential of the community. Several new approaches, such as stable isotope probing (SIP) and multi-isotope imaging mass spectrometry (MIMS), have been developed to study structure–function relationships of microbial communities. The combination of biogeochemical stimulation experiments with these functional and/or molecular approaches has the potential to elucidate the role of specific microorganisms concerning specific biogeochemical functions. These experimental analyses, together with *in situ* measurements of molecular and biogeochemical markers, enable the identification of key catalysts of pelagic biogeochemical processes and determination of their regulatory factors. Based on our current understanding of the structure–function links in bacterioplankton, we hypothesize that any local bacterial community comprises 2 parts: the abundant members form a core community that carries out the ongoing biogeochemical functions, and the rare members form a seedbank that provides the genetic potential to respond to any change in the system. Together, both parts form a pan community that is able to mediate the biogeochemical processes in the pelagic ecosystems of the earth.

KEY WORDS: Community structure · Pyrosequencing · Metagenomics · Bioinformatics · Comparative genomics · Structure–function links · Drivers · Community theory

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

An enormous effort has been made during the last 20 yr to assess the diversity of natural microbial communities using molecular methods (Schmidt 2006). One reason for this effort was the discovery that most microbes from natural environments defy cultivation by conventional methods and are therefore only amenable to cultivation-independent molecular

approaches. Another reason was to reach a comparable level of understanding of the community structure in microbial ecology as for higher organisms, i.e. identify the species present in a given sample and determine their abundance. For aquatic ecology, this understanding is especially relevant because most biogeochemical processes in aquatic ecosystems are controlled by microorganisms. Another prime reason for the molecular assessment of microbial diversity is

*Email: mho@gbf.de

the belief that knowing all the microbial members of a natural community would be a major step towards understanding and predicting its biogeochemical functioning. Prediction is one of the overall aims of science, and predictive ecology is a well-developed ecological discipline. Biology is lacking predictive capacity in some areas, such as biogeography and taxonomy, because the prime goals of these fields are to describe nature and to develop ontologies. However, the predictive power of science is needed to answer many questions society is asking in order for a better and more sustainable use of the environment.

The pelagic environment of aquatic ecosystems is certainly one of the largest habitats on earth, and its prokaryotic community represents the biggest biomass and largest reactive surface in this environment. In addition, open water covers more than 70% of the earth's surface, providing a highly reactive interface to the atmosphere, and controls major parts of the earth system including the climate. In this article, we concentrate on bacterioplankton as an example of a natural microbial community of global relevance. We review and comment on the state of knowledge of the diversity of bacterioplankton, the different contributions made by the primary approaches taken to gain this knowledge, and future research directions necessary for a predictive biogeochemistry of pelagic environments.

CURRENT UNDERSTANDING OF MICROBIAL DIVERSITY IN PELAGIC ECOSYSTEMS

Two trends characterized the development of molecular approaches during the last decade: (1) efforts to combine different approaches in order to identify specific metabolic functions of specific microorganisms and (2) the assessment of all genes present in a given habitat, i.e. the metagenome of a microbial community. The last few years have seen the advent of new sequencing technologies (Margulies et al. 2005, Hall 2007) that generate orders of magnitude more sequence data per run than previously obtained. It is now possible to sequence the genome of a single microbial cell (Marcy et al. 2007).

In microbial ecology, the term species is still under debate because different methods result in different definitions for different groups of bacteria (Cohan 2002, Gevers et al. 2005). Our current understanding of a microbial species is based on the molecular analysis of

DNA sequences from single strains or sets of strains up to the sequence of whole genomes. The 3 main molecular methods currently used—16S rRNA gene sequence analysis, multilocus sequence typing (MLST) and multiple locus variable number of tandem repeat analysis (MLVA)—identify bacterial strains with different taxonomic resolution (Fig. 1). The gold standard of microbial taxonomy is 16S rRNA gene sequence analysis followed by MLST, which uses a set of housekeeping gene sequences (usually 7). MLVA achieves the highest resolution by using the rapidly evolving variable number of tandem repeats (VNTR) in a fingerprint analysis, but it does not provide genetic information for the determination of the phylogenetic relationship among strains (Lindstedt 2005). Both high resolution techniques have only been applied to cultured bacteria and not to natural microbial communities.

All bacterial species definitions suffer from having arbitrary cut offs, e.g. 97% similarity for 16S rRNA sequences or 70% for DNA–DNA similarity, which are not valid for all taxa. In addition, it is not clear what taxonomic resolution is needed to explain a specific biogeochemical function. Considering all these problems with species definition, we first concentrate on the 16S rRNA as a taxonomic yard stick. In microbial ecology, it is important for every assessment of microbial diversity to define an operational taxonomic unit (OTU) that specifies the procedure by which this unit is determined. The taxonomic level of the OTU could, in principle, range from the domain to the subspecies level, but for comparative purposes it is very important to have the same level of taxonomic resolution and not, for example, to compare classes with genera. The appropriate phylogenetic resolution might depend on

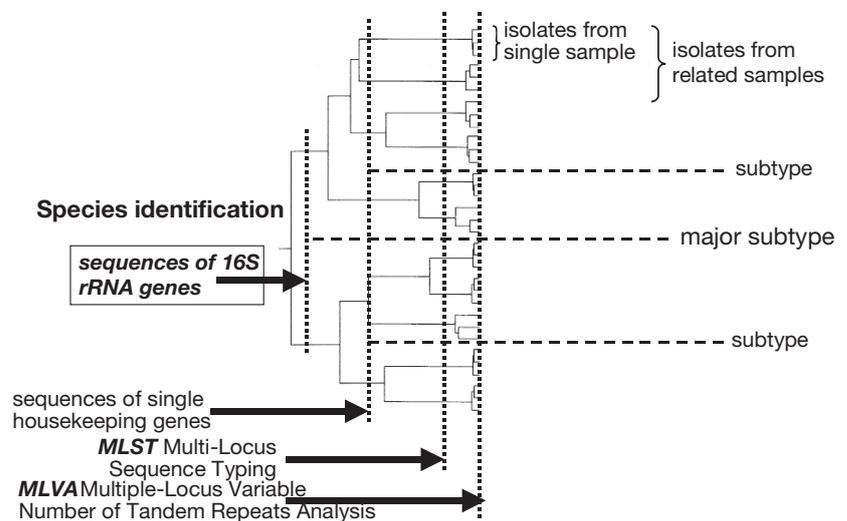


Fig. 1. Taxonomic resolution of different DNA-based genotyping methods for bacterial strains and isolates

the specific question asked. For example, in the assessment of the pathogenic potential of a water sample contaminated with *Vibrio cholerae*, the serotype (subspecies) is of relevance to human health, whereas for the understanding of ammonium oxidation in the water column, it would be important to know which phylum the ammonium oxidizers belong to. Considering the fact that almost all our detailed knowledge about the physiology and biochemistry of bacteria stems from pure culture studies, it might be most valuable to aim at a species-level resolution in the attempt to link community structure to biogeochemical function of aquatic communities. However, which level of taxonomic resolution has to be addressed might depend on the respective questions and the methodology used.

Absolute abundance, i.e. determination of the number of cells of a specific species per volume unit, is normally estimated using single cell detection methods such as fluorescent *in situ* hybridization (FISH) or immunofluorescence microscopy. In the best case, the detection limit of immunofluorescence microscopy is in the range of a few cells ml⁻¹ (Brettar & Höfle 1992). For FISH, a relative detection limit of 1% is generally assumed (Amann et al. 1995). On the other hand, it is not practical to use these single cell approaches to determine the abundance of all species in a natural sample. This has been attempted using clone libraries, and fingerprints coupled with the sequencing of single clones and bands excised from the fingerprints, respectively. Both approaches generate only relative abundance data at best. There is an ongoing discussion about the insight into bacterial communities provided by polymerase chain reaction (PCR)-based clone libraries and fingerprints. Provided that both are generated by PCR amplicons using the same primers, the results should be comparable as a first approximation. Fingerprints provide an overview on the abundance of the PCR amplicons down to a detection level of, usually, 0.1% of the total amplicons (Pedrós-Alió 2006). Provided that, on average, 10¹¹ amplicons are loaded onto a fingerprint lane—assuming 30 ng of amplicon on a single strand conformation polymorphism (SSCP) fingerprint (Brettar et al. 2006)—every prokaryotic cell in a sample of 10⁹ cells l⁻¹ (1 l surface water extracted) is on average represented by 100 amplicons, providing the basis for a high-resolution estimate of the abundance of OTUs above an average visibility threshold of about 10⁸ ampli-

cons (theoretically corresponding to an abundance of 10³ cells ml⁻¹). To achieve a similar estimate of the abundance of clones in a library, more than 1000 clones would have to be screened. The advantage of clone analysis might be more 16S rRNA sequence data, providing a better phylogenetic resolution. Fingerprints of rare amplicons may be hidden by the strong bands of abundant amplicons. However, there is only a likely chance of detecting rare clones by clone screening when a large number of clones is analyzed. The frequent lack of comparability between fingerprints and clones may thus result from an insufficiently low number of screened clones, resulting in different levels of resolution. This, in turn, leads to a bias in the estimated abundance of clones compared to the fingerprint. However, since today's sequence technologies facilitate large-scale clone screening, we think that the results from clone libraries on one side and fingerprints on the other will agree more frequently in the future.

These considerations are also relevant for the assessment of species richness, because any assessment of richness determines only the number of species (or OTUs) above the detection limit of the method used. To portray evenness, rank-abundance curves are quite useful, but have rarely been used in aquatic microbial ecology until now. Rank-abundance curves usually show a log normal distribution (Sloan et al. 2006) and have the potential to discriminate between rare and abundant species. Fig. 2 shows rank abundance curves of SSCP community fingerprints of 2 bacterioplankton communities from very different habitats in the central

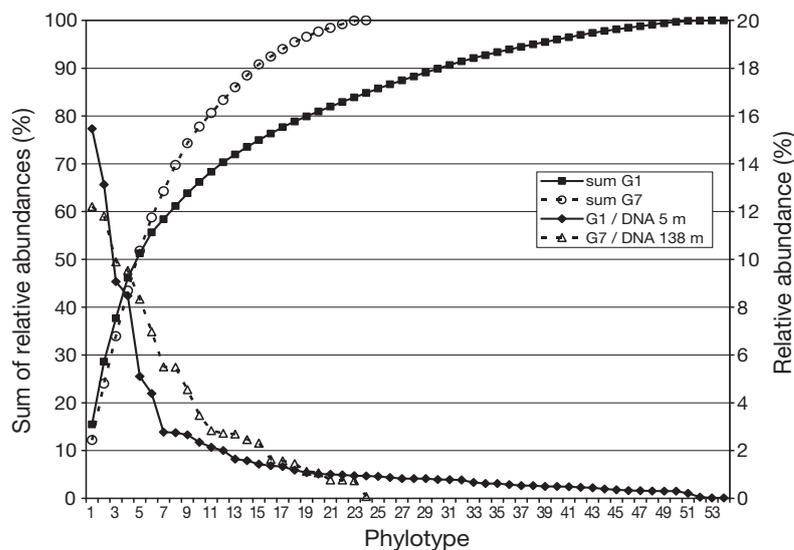


Fig. 2. Comparison of rank-abundance curves of phylotypes in 2 bacterioplankton samples from different water depths (sample G1 from 5 m, aerobic, and sample G7 from 138 m, anaerobic) from the central Baltic Sea. Data are based on the relative abundances of phylotypes from 16S rDNA community fingerprints (Brettar et al. 2006) using single strand conformation polymorphism analysis

Baltic Sea, i.e. the surface and the anaerobic deep water (Brettar et al. 2006). These rank-abundance curves show 2 different branches with an inflection point around 2 to 3% of relative abundance of single phylotypes in the community fingerprint. About half of the detected phylotypes of the anaerobic bacterioplankton are abundant, whereas only about 14% (7 out of 51) of the phylotypes are abundant in the surface bacterioplankton community if the inflection point is used as a threshold criterion (Fig. 2). The richness of the community is best seen if the cumulative sums of the relative abundances are calculated. The richness determined from these cumulative curves indicates 21 OTUs (phylotypes) for the anaerobic community and 51 OTUs for the surface community, assuming a threshold of 0.1%, which is the common detection limit for SSCP community fingerprints (Pedrós-Alió 2006). This means that more than 99% of the cells in the anaerobic community were represented by about 20 phylotypes, whereas the aerobic community had more than twice as many members in this fraction. The curves also show that the large majority of the surface bacterioplankton are rare phylotypes with abundances between 2 and 0.5% of the community. The general finding of a few abundant and many rare OTUs is also in accordance with clone library studies of bacterioplankton communities (Pommier et al. 2007).

How real these percentages are needs to be determined, because fingerprints or clone libraries are PCR-biased (von Wintzingerode et al. 1997). One approach to validating these relative abundances is quantitative PCR, using the sequence information obtained from the community fingerprints or clone libraries (Labrenz et al. 2004, Campbell et al. 2008). Overall, it is important to elucidate the 'true' community structure by determining the abundances of both the abundant and the rare members of the community at about the species level in order to link diversity and function in microbial communities and understand the microbiologically mediated biogeochemical processes in the water column.

ACCESSING THE RARE BIOSPHERE BY PYROSEQUENCING

One promising novel approach to reaching the ideal of the 'all-abundant-species' assessment for bacterioplankton could be pyrosequencing, as demonstrated recently (Edwards et al. 2006, Roesch et al. 2007). This approach has been applied to the V6 region of the 16S

rRNA of bacterioplankton (Sogin et al. 2006, Huber et al. 2007). In the most detailed analysis of a single bacterioplankton sample, more than 440 000 sequence tags were obtained (Huber et al. 2007). This approach provides the information needed to describe the bacterial community structure, i.e. the OTU identity by the sequence of the tag, and the abundance of the OTU by the abundance of the tags. The most abundant tag was found to be present more than 70 000 times, i.e. with a relative abundance of 29% of all tags. While the abundances of the tags dropped rather rapidly, the first 10 most abundant tags comprised about 55% of all tags analyzed (Fig. 3).

The real power of this approach is in the analysis of the rare members of the community, named 'rare biosphere' by Sogin et al (2006), which comprised about one-third of all tags. Fig. 3 shows that the number of these very rare members increases enormously: as OTUs become more rare, more single tags are observed. In the end, at an abundance of one tag, more than 12 000 tags were detected. This means that there is still a large number of phylotypes observed at a detection limit of 0.0002%. The next important question is: What is the phylogenetic resolution of the approach—in this case, the phylotype defined as a sequence tag of an average length of about 60 nt in the V6 region of the 16S rRNA gene? To assess the phylogenetic resolution of the tags in the Huber et al. (2007) dataset, we compared all unique tags from both bacterioplankton samples (about 30 000 tags) and found a relatively similar taxonomic resolution in both samples (Fig. 4). For this comparison we used only the tags that

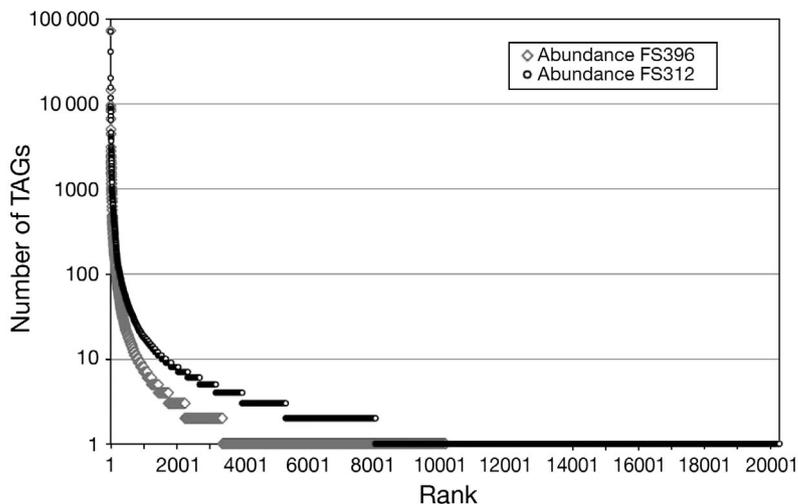


Fig. 3. Absolute rank-abundance curves of 2 bacterioplankton samples (FS396 and FS312) analyzed using pyrosequencing of the V6 region of the 16S rRNA, recalculated from original data released by Huber et al. (2007) at jbpc.mbl.edu/research_supplements/g454/20070822-private/supplemental.zip, which contains all the fasta-formatted trimmed reads used in the analyses (J. A. Huber pers. comm.)

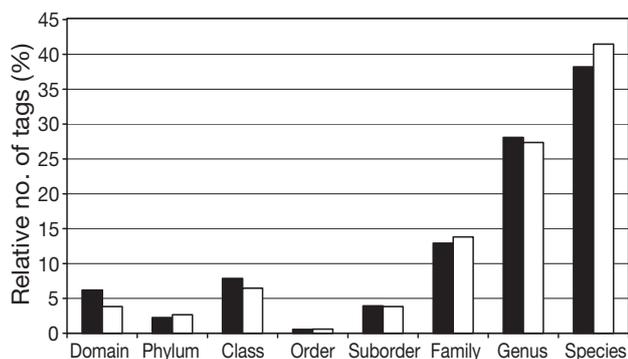


Fig. 4. Taxonomic resolution of the sequence tags from Huber et al. (2007). The relative number of tags that could be identified at the given taxonomic level for the 2 samples (empty boxes: FS 312, filled boxes: FS 396) is shown. Less specificity is expected when more cultured species are sequenced (16S rRNA gene sequences from only 8000 species were used)

had a cultured counterpart, because for cultured bacteria, the taxonomic position is better defined. Fig. 4 shows that about 40% of such tags can be assigned to a distinct species and about 28% to a genus. The rest of the tags, i.e. about one-third of all tags, can only identify the family or even higher taxonomic ranks. This looks quite good from an ecological point of view, although some analytical problems remain (Liu et al. 2007).

The power of the pyrosequencing approach lies in its capacity to generate massive amounts of sequence data for relatively low cost. However, its suitability for linking function and community structure seems limited. The shortness (currently about 200 bp) of the sequence reads poses a challenge in assembling contigs for natural microbial communities because of their high diversity, but bioinformatic solutions are conceivable (Liu et al. 2007, Sundquist et al. 2007). The strength of pyrosequencing for community analysis lies in increasing knowledge about the biogeographical distribution of bacteria, since it can provide detailed information on the relative abundance of individual phylotypes, including those with very low abundances. This technique offers the possibility to address ecologically relevant questions such as the role of the high numbers of low-abundance phylotypes in natural environments (Pedrós-Alió 2006). Thus far, pyrosequencing has only been used once in experimental systems in which the microbial diversity was altered (Mou et al. 2008). Such experimental studies on the diversity–productivity relation, including the recently developed *in situ* expression profiling (Frias-Lopez et al. 2008), or on the diversity–stability, seem to be particularly suitable for pyrosequencing and could contribute to a better incorporation of microbes in community theory.

METAGENOMICS AND STRUCTURE–FUNCTION LINKS

Until now, we have only looked at the bacterial community structure using the 16S rRNA gene. To understand and predict the biogeochemical functioning of the bacterioplankton community, we have to look at functional genes present. Gene content and the regulatory networks governing gene expression are well defined in the genome sequence of a bacterium. To date, more than 400 bacterial genomes have been sequenced and analyzed with several hundred genomes in the pipeline. The genomes of many of the most abundant aquatic bacteria, like *Prochlorococcus marinus* and *Pelagibacter ubique*, are fully annotated and analyzed in detail (Dufresne et al. 2003, Giovannoni et al. 2005). Therefore, a promising approach for linking community structure of prokaryotes with biogeochemical processes—the structure–function question—could be metagenomics or environmental genomics (DeLong 2005).

Both terms refer to the use of genomic approaches to examine microbial genomes in natural environments without isolating the microbes into pure cultures. The term ‘metagenomics’ emphasizes the fact that this approach examines several microbes and their genomes within a single sample (Handelsman 2004). Like genomics applied to a microbe in pure culture, metagenomics retrieves several genes (or at least their sequence) simultaneously, even though only a few (e.g. proteorhodopsin and 16S rRNA genes) are of immediate interest. For example, a metagenomic study may use PCR to screen a fosmid library (see below) for proteorhodopsin and 16S rRNA genes, but examining clone libraries of proteorhodopsin and 16S rRNA genes retrieved by PCR would not be a metagenomic study.

Here, we introduce some basics of metagenomics in order to illustrate the power and the weaknesses of this approach with respect to examining the structure–function question. We focus on using metagenomic sequence data to assess both potential biogeochemical roles and the phylogenetic position of uncultivated microbes. Space limitations prevent us from discussing other metagenomic approaches, such as microarrays (He et al. 2007), for exploring structure–function relationships among aquatic microbes.

There are 4 general approaches for obtaining metagenomic sequence data (Fig. 5). Perhaps the easiest technically is to clone relatively small fragments of genomic DNA (‘small inserts’) into plasmid or phage vectors. The resulting clone library can then be screened by probing for genes of interest (Schmidt et al. 1991), by looking for enzyme activity (Cottrell et al. 1999), or by sequencing the ends of the insert (Venter et al. 2004). Usually, these clones yield sequence data

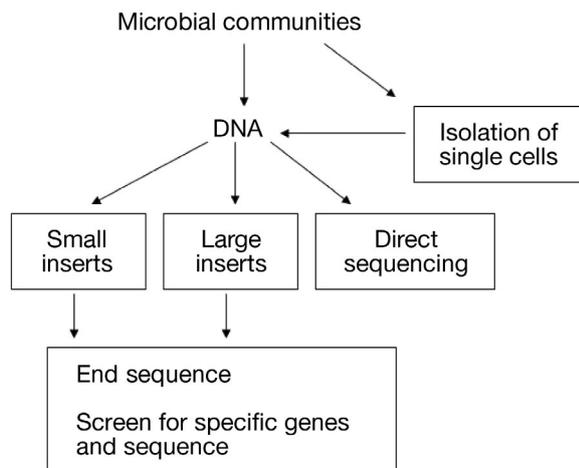


Fig. 5. Summary of the 4 types of metagenomic approaches. 'Small' and 'large' inserts refer to the size of the foreign DNA cloned in metagenomic libraries. The dividing line varies with the cloning vector, but is roughly about 5 kbp. 'Direct sequencing' refers to the application of new sequencing technologies, such as pyrosequencing. The 4th approach, isolation of single cells, usually depends on separating cells by flow cytometry

from only 1 or 2 genes. The small insert library approach, not the PCR-based method that is now standard, was used by one of the first studies examining 16S rRNA genes in the ocean (Schmidt et al. 1991), long before the advent of even pure culture genomics. Small insert clone libraries have been used by some of the largest and most extensive metagenomic studies, including that of the Sargasso Sea microbial community (Venter et al. 2004) and the Global Ocean Survey (Rusch et al. 2007).

Sequence data from small insert clone libraries are especially useful for linking structure and function when sequences are used to make contigs. These are constructed *in silico* with sequences from 2 or more clones that overlap (e.g. 40 base pairs that are 94% identical, which is the default setting for the Celera assembler used by Venter et al. 2004, J. Heidelberg pers. comm.) and thus share highly similar sequences in the overlapping region. Contig data can be brought to bear on the structure–function question when the contigs have a phylogenetic marker (the prized one being the 16S rRNA gene) and a gene (or several) encoding enzymes for a biogeochemically relevant function. One important example is the contig in the Sargasso Sea dataset with a 16S rRNA gene from an archaeon and a gene for ammonium monooxygenase, which catalyzes the first step in ammonium oxidation (Venter et al. 2004). However, only a relatively small fraction of the entire dataset can be used for examining the structure–function question, mainly because of the high diversity of microbial communities. Of nearly 2

million sequence runs in the Sargasso Sea dataset, only about 1 210 000 contigs of any size can be constructed (Venter et al. 2004); of these, only 4262 are larger than 5 kbp and possibly bear more than a couple of genes (M. T. Cottrell pers. comm.). More problematic, it is possible that the contigs constructed *in silico* may be artifacts and may not be gene fragments from one individual microbe.

The second approach is probably more useful for linking prokaryotic community structure with specific biogeochemical processes (Fig. 5). This approach starts with the cloning of large fragments of microbial DNA into vectors that can accept about 40 kbp and over 100 kbp of foreign DNA in the case of fosmids and bacterial artificial chromosomes (BACs), respectively. Large insert libraries can be screened with the same methods used to screen libraries with small insert clones. This approach can be difficult because of the need to isolate and work with high molecular weight DNA. However, one is assured that 2 or more genes found on a large insert clone are from one individual microbe. The presence of a phylogenetic marker and a biogeochemically relevant gene on a single fosmid or BAC clone establishes one link between structure and function for the analyzed microbial community.

A good example of the power of this metagenomic approach is the discovery of proteorhodopsin in oceanic surface waters (Béjà et al. 2000). This form of potential photoheterotrophy was revealed by sequencing a BAC clone, which was selected for detailed analysis after the clone library was screened for 16S rRNA genes. The sequence data indicated that the clone possessed a 16S rRNA gene from a member of the *Gammaproteobacteria* (SAR86) and, serendipitously, also a gene similar to an archaeal rhodopsin (misnamed 'bacteriorhodopsin'), which was confirmed to function as a light-driven proton pump by protein expression experiments (Béjà et al. 2000). Subsequently, small insert clone libraries revealed a high diversity of proteorhodopsin genes in the Sargasso Sea, suggesting that many bacteria other than just the SAR86 clade have proteorhodopsin (Venter et al. 2004). However, it was again data from a contig that made the link between this gene and another bacterial group, the *Bacteroidetes* (Venter et al. 2004).

The third metagenomic approach does not rely on clone libraries; rather, it directly sequences genomic DNA isolated from uncultivated microbial communities (Fig. 5). This approach is now feasible due to the development of new sequencing technologies that can generate massive amounts of sequence data for relatively low costs (Margulies et al. 2005). So far, one new sequencing approach (pyrosequencing) has been used for examining the diversity of 16S rRNA genes retrieved by PCR (see 'Accessing the rare biosphere by

pyrosequencing' above). In theory, these new technologies could be applied directly to microbial DNA or RNA without PCR.

The fourth and final metagenomic approach reduces the diversity of the sample by examining the genomes of single cells usually isolated by flow cytometry (Fig. 5). It could be argued that an analysis of a single cell is not a 'metagenomic' study because only one genome is examined, but it remains part of environmental genomics because it avoids cultivation of the microbe. In any case, and in part because of the many technical difficulties, this approach has so far rarely been used with marine microbes. The feasibility of this approach, albeit with pure cultures, was explored for the cyanobacterium *Prochlorococcus* (Zhang et al. 2006). This microbe, which is the most abundant photoautotroph in nature, is easily identified and thus separated from other microbes by flow cytometry because of its unique autofluorescing pigments (Chisholm et al. 1988). Another study flow-sorted non-pigmented prokaryotes from coastal waters and randomly selected a few cells for genomic analyses (Stepanauskas & Sieracki 2007). Of the 11 single amplified genomes (called 'SAGs' by the authors) analyzed, 5 were flavobacteria (*Bacteroidetes*), and 2 of these bore rhodopsin genes. The genomes of single cells identified using FISH have been partially sequenced (Podar et al. 2007), but flow cytometric sorting of FISH-labeled microbes remains difficult and far from routine. Even so, single cell genomics is potentially a very powerful approach for exploring structure–function questions in aquatic microbial ecology.

The limitations of an approach are as important to examine as the advantages, and metagenomics is no exception. Many of these limitations are shared by genomics of microbes in pure cultures. Recognizing a stretch of DNA sequence as a gene is still difficult (Huson et al. 2007), and the function of many genes—as many as half—remains unknown even for well-studied microbes. Identification of the function of a gene by sequence analysis alone can be misleading (Cottrell et al. 2005b). Few studies have examined whether metagenomic libraries adequately represent the genetic composition of natural microbial communities (Cottrell et al. 2005a), and there is some evidence that, for example, the SAR11 clade may be underrepresented in fosmid libraries (Campbell et al. 2008).

Perhaps the most serious limitation is that the presence of a gene does not necessarily mean that the biogeochemical function mediated by that gene is being carried out in the environment under study. This question is related to a long-standing problem in microbial ecology: the presence of a microbe does not necessarily mean that it is active, or more succinctly, abundance does not equal activity. However, there are examples

where abundance of a particular bacterial group is a good predictor of its contribution to a biogeochemical function, such as biomass production as measured by leucine incorporation in the example given in Fig. 6. While there are certainly examples to the contrary, a reasonable null hypothesis is still that activity does in fact follow cell abundance and the presence of a gene. The presence of a gene or organisms and microbial activity seem likely to be linked because of a strong selection against inactive cells (they may be eaten) and unused genes. The pattern of activity might most strongly change if we move down the water column, into the large ocean basins.

Regardless, metagenomic studies can, of course, not be used to conclude anything definitive about microbial activity in aquatic systems (Oremland et al. 2005). What is more interesting to note is that metagenomic studies have suggested types of microbial activity never before suspected to occur in the surface layer of aquatic systems. Other examples in addition to proteorhodopsin include genes related to the use of carbon monoxide and reduced sulfur in the surface ocean (Moran et al. 2004, Moran & Miller 2007). However, the *in situ* activity of many of these genes remains to be demonstrated. Similar to genomic studies of organisms in pure cultures, metagenomics may be more important in generating hypotheses than in confirming them.

CURRENT LIMITATIONS OF BIOINFORMATIC ANALYSIS

Several issues in bioinformatics are relevant for analyzing microbial communities and their biogeochemical functions. A major issue is the increasing amount of data. 'Normal' public databases, such as the DNA Data-

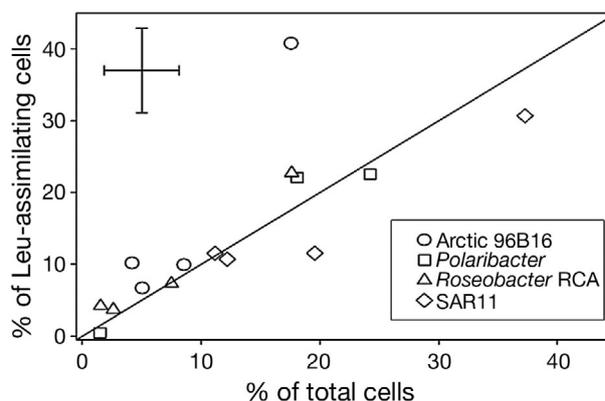


Fig. 6. Contribution of 4 bacterial groups to leucine (Leu) assimilation (% of all cells active in Leu assimilation) versus their abundance. Reproduced from Malmstrom et al. (2007). Error bars in the upper left are average SE calculated for all points

bank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL) and GenBank, were designed to store and publish manually obtained sequences, using for example the Sanger method (Sanger et al. 1997). Until 2004, the amount of data published roughly doubled every year, but with the advent of new technologies, this rate is likely to explode (Fig. 7). Until 2006, it was rather easy to use a simple web browser to access a public database and to use search engines such as ACNUC, Entrez or SRS to retrieve sequences by keywords, or to use tools such as BLAST (www.ncbi.nlm.nih.gov/blast/Blast.cgi) to retrieve sequences by similarity. Times are changing, and for a biologist, there are now simply too many sequences deposited in the public databases than can easily be dealt with. Aside from regular sequencing, the time of 'global sequencing' has arrived, and we are probably not yet ready to cope with the mass of data generated by these types of projects, which are, in terms of sequence information, in the range of the human genome (Rusch et al. 2007).

Below, we look at available databases for 'standard sequences', metagenomic data and new pyrosequencing data. As for standard sequences, almost any DNA sequence potentially coding for a protein (open reading frame, ORF) is translated and deposited in the public database of proteins. UniProt (4 949 164 entries, www.uniprot.org), for example, is divided into 2 divisions: SwissProt (276 256 entries), which contains proteins duly identified and well annotated by experts, and Trembl, which contains most of the remaining entries. Half of these proteins originate from bacteria, and by the end of 2007, one-third of the bacterial pro-

teins included in the database had been deposited within the previous year (Fig. 7). UniProt only doubles in size every year, because it does not include metagenome data. In contrast, GenBank (update December 11, 2007) contained 1 140 983 ribosomal sequences, 621 900 being annotated as 16S rRNA gene sequences. Note that these 16S rRNA gene sequences were mostly short to very short (50 to 500 nt, 248 972 entries), and only 201 076 entries had a length of 1200 nt or more, of which half were submitted in 2007 (Fig. 7). Only 32 880 entries of these long sequences belong to cultured strains comprising about 8000 different species.

The UniProt Metagenomic and Environmental Sequences (UniMES, introduced on May 29, 2007) database is a repository specifically developed for metagenomic and environmental data, which currently contains only data from the Global Ocean Sampling (GOS) Expedition (Rusch et al. 2007, source EBI, November 21, 2007). These 6 028 191 sequences can be retrieved using keywords with Entrez, or using a BLAST query on the 'Environmental samples' database at the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov/). In EMBL/GenBank/DDBJ, the 'env_nt' division (nucleotide sequences from environmental samples) includes data from projects such as the Sargasso Sea and mine drainage (Tyson et al. 2004, Venter et al. 2004). As a result, these earlier datasets can be queried using keywords, and BLASTed the usual way. Each sequence is deposited under a separate entry with its own accession number. Each sequence may also be annotated for

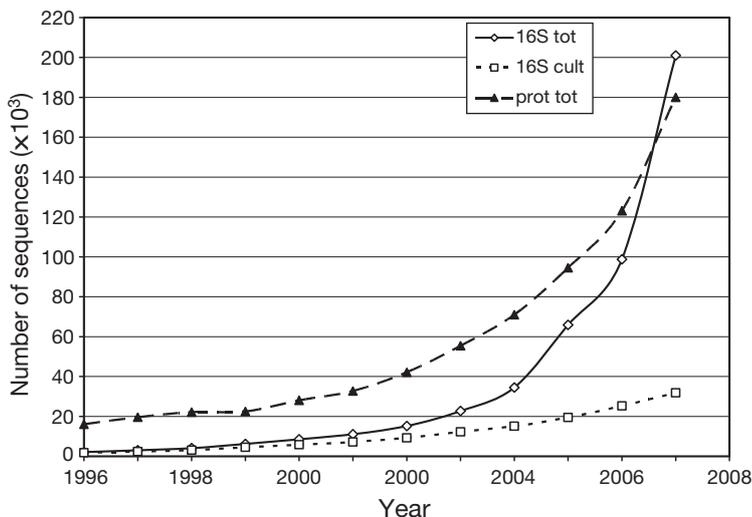


Fig. 7. Increase in the number of bacterial 16S rRNA and protein sequences deposited in international databases over the last 10 yr (status end of 2007). Only sequences of at least 1200 nt were considered for 16S rRNA. The total number of 16S rRNA gene sequences (\diamond) also includes those from cultured bacteria (\square). Bacterial protein sequences are from SwissProt (well annotated sequences) and do not include data from Trembl or UniMes

sequence similarity, product, etc. In early 2001, the NCBI and Ensembl (<http://trace.ensembl.org/>) developed the Trace Archive. This has, for some time, successfully served as a repository for new metagenomic sequences produced by the traditional Sanger-based sequencing protocols. As of October 2007, the Trace Archive contained 3760 projects (folders) that can be accessed at <ftp://ftp.ncbi.nih.gov/> (folders 'pub' then 'TraceDB'). Sequences have no accession number; they can be BLASTed (on these servers) or searched, but the retrieval tools are relatively crude. Finally, it is worth mentioning that data are often also available for download from author's web sites, and most can be obtained from the Camera web site (<http://camera.calit2.net>), but without programmatic access.

Microbiologists are currently sequencing every tag (usually in the 60 to 80 bp range) obtained using PCR amplification of a short domain of the 16S rRNA gene with the help

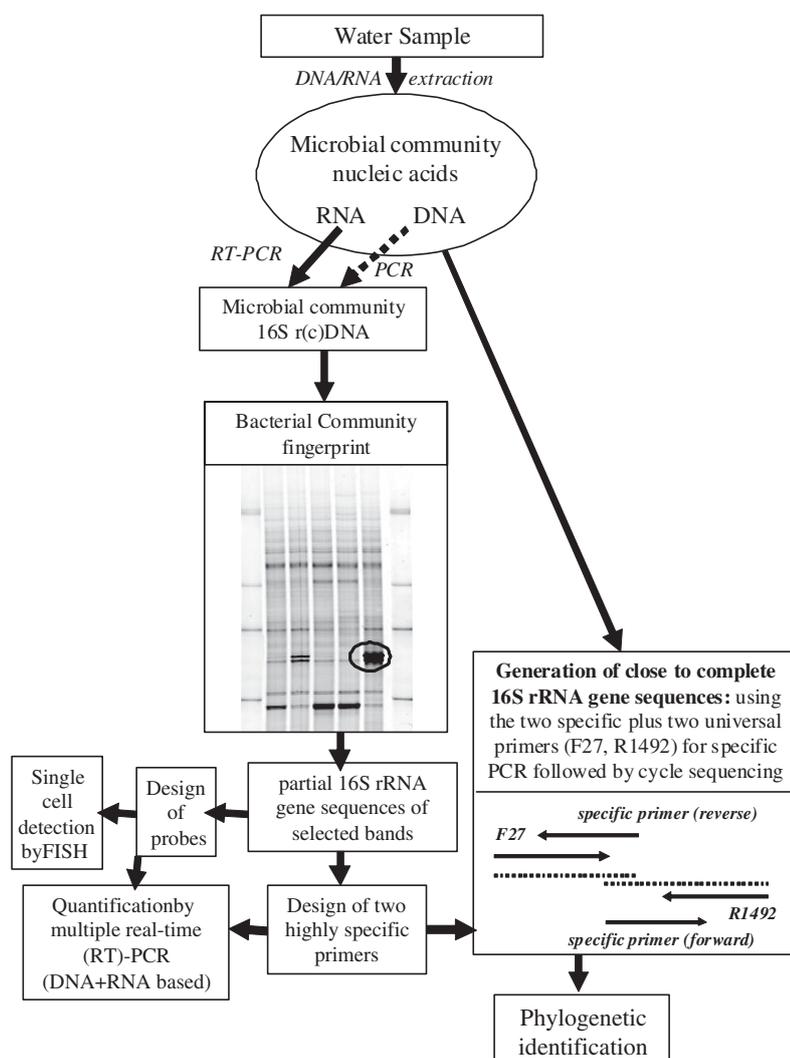


Fig. 8. Analysis of *in situ* samples and experiments for identification of bacterial key catalysts in a fingerprint-based approach

of pyrosequencing technology. Due to the structure and volume of these pyrosequencing data, they cannot fit into the current Trace Archive design. A new archive was therefore built in October 2007, the Short Read Archive (www.ncbi.nlm.nih.gov/Traces/sra/). In this database, data are in new specific formats, which may be difficult to read for most users. In addition, the NCBI Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo/) now accepts pyrosequencing data, although it is probably for expression data only. Retrieval of pyrosequencing datasets is not always easy. For example, some sequences from the same sample may be in the DDBJ/EMBL/GenBank databases while others, being too short, are not. The Camera web site (<http://camera.calit2.net/>) was purposely designed to store environmental data, but pyrosequencing data is

not included and the unpractical access to this site (manual login required) is a real problem for bioinformatics.

Metagenomic and 16S rRNA gene pyrosequencing analyses are 2 very different cases. Metagenomic data focus more on getting a complete gene inventory of an environment and reconstructing full genomes if possible. The goal of metagenomics comprises both analyses of diversity and identification of genes potentially relevant to biogeochemical cycles. As a result, analyses of these datasets are extremely difficult and often require much computing power and a team of scientists from different fields for interpretation, while mainly focussing on a single sample. In contrast, 16S rRNA gene pyrosequencing data are fairly homogeneous and easier to analyze, as the only goal is to study diversity. Because sequences are short (tags), general tools such as BLAST and alignment programs should be replaced by ad hoc new programs. Comparisons of different datasets are still missing.

Finally, in contrast to the analysis of complete genomes, environmental bioinformatics lack dedicated servers which could provide the following: (1) dedicated new tools, as they become available, either as programs for download or to be run from the web server, (2) manually curated databases, similar to Swiss-Prot for proteins, for 16S rRNA gene sequences and interesting house-keeping genes, and (3) storage facilities for intermediate analyses that allow comparison across single experiments or

analyses. In conclusion, as the scientific community organizes itself to deal with the present deluge of sequences and as more complete genomes from aquatic microorganisms become available, analyses will become more powerful, provided that dedicated knowledge databases, such as ontologies, are also developed by microbiologists.

INSIGHTS FROM COMPARATIVE GENOMICS AND THE PHYSIOLOGY OF CULTURED BACTERIOPLANKTON

The study of environmental isolates is attracting new interest due to the need for habitat-specific genome information, and based on several success stories, like

the isolation of aquatic *Actinobacteria* (Hahn et al. 2003) and representatives of the SAR11 clade (Rappé et al. 2002, Giovannoni & Stingl 2007). This development could solve one of the key questions for linking microbial diversity with biogeochemical function: which level of taxonomic resolution has to be achieved to obtain enough information for biogeochemical predictions? Some insight into this question can be gained when comparing species with a larger set of strains obtained at the same habitat but at different times, such as in the case of *Shewanella baltica* (Ziemke et al. 1998, Höfle et al. 2000) and *Rheinheimera baltica* (Brettar et al. 2002a). Both species were obtained from the same station in the central Baltic Sea. While *S. baltica*, a rapidly growing and highly versatile r-strategist, displays a high physiological diversity on the sub-species level, the slow-growing strains of *R. baltica* have very similar physiological features. To analyze the function and the ecological niches of bacteria with sub-species diversity, as in the case of *S. baltica*, even the clonal level seems to be relevant (Ziemke et al. 1997), while the species level could be considered as sufficient for bacteria like *R. baltica*. The genus level seems to be too rough in many cases. For example, the physiology of *Shewanella denitrificans* (Brettar et al. 2002b) is so different from the closely related *S. baltica*, its neighbor in low oxidic waters of the Baltic Sea, that a completely different niche has to be expected and was indeed observed (I. Brettar pers. obs.). These results, gained by physiological studies and clonal analysis of aquatic isolates, were recently confirmed by comparative genome analysis of 4 *S. baltica* and 1 *S. denitrificans* strain (strains OS155, OS185, OS195, OS223, OS217, see http://genome.jgi-psf.org/mic_home.html).

Comparative genomics have brought a new perspective to the attempt by Vandamme et al. (1996) to define a bacterial species as 'the condensed nodes in a cloudy and confluent taxonomic space'. Several authors now advocate basing species definition on genome analysis (Coenye et al. 2005, Konstantinidis & Tiedje 2005). A major challenge for microbial ecologists is the different degree of variability of the genome encountered within a single species. The degree of difference between the core genome and the strain-specific genes can vary greatly from species to species. For example, some fast-growing, versatile bacteria like *Escherichia coli* and *Streptococcus* B group seem to have a rather large pan genome (the total of all genes of a species) (Medini et al. 2005). There is evidence that this is also the case for the fast-growing and highly versatile *Shewanella baltica* (see genome sequences of OS155, OS185, OS195, OS223, with genome sizes of 5.1 to 5.4 Mb) (Höfle et al. 2000, Deng et al. 2008). In contrast, the slow-growing SAR11 representative *Pelagibacter ubique* has a small, highly conserved

genome with only 4 small highly variable parts, as shown recently for strains of *P. ubique* as well as by metagenomic data analyses of the Sargasso Sea (Wilhelm et al. 2007).

Currently, the following picture emerges if we use the r–k continuum concept for ecological strategies (Hirsch 1979): the available genomes show the tendency of a larger ratio of pan genome to core genome for fast-growing and versatile r-strategists compared to slow-growing and resource-efficient k-strategists. *Pelagibacter ubique* and *Prochlorococcus* are prime examples for aquatic k-strategists, both having a small, streamlined core genome with limited variation among strains. However, the genome-shaping forces are not well understood at present, and our understanding of what is a species and what is an ecotype will greatly profit from the outcome of comparative genomics. In addition, the impact of intraspecies and interspecies competition and 'collaboration' could be reflected by the genome or the pan genome and the underlying genome-shaping forces.

Results of comparative physiology and genomics of marine bacterial strains are consistent with the existing ecotype concept, in which an ecotype covers its ecological niche with specific features and functions and can be phylogenetically regarded as a set of genetically cohesive clones of a specific species (Gevers et al. 2005). However, as shown by the analysis of the clonal structure of *Shewanella baltica* over time and space, and now supported by comparative genomics, it is a mixture of ecotypes with various features and the respective encoding genes that finally seem to drive the evolutionary process as a 'functional and interacting unit', i.e. species. From this perspective, we conclude that the species level is usually sufficient in order to link the organism to the process. However, there are certainly specific questions that require resolution at the subspecies/ecotype level in order to be addressed.

Concerning intraspecies diversity, i.e. the clonal structure of a species, we have to take into account that, from an ecological point of view, there is a great difference whether intraspecies diversity exists in the same sample or whether the ecotypes of a species are separated by time and space. For example, many ecotypes of *Shewanella baltica* exist even within the same liter of a water sample (Höfle et al. 2000). In this case, many interactions, such as the exchange of genes and/or substrates, are likely. These types of interactions can be excluded if the studied ecotypes are separated by space (e.g. different habitats like different depths or geographically different regions) or time (e.g. ecotypes occurring only in winter, and others only in summer). This means that, in the case of high intraspecies diversity in a sample, intraspecies interaction will have a different impact than in the case of low intraspecies diversity. Therefore, the con-

comitant presence of a high or a low number of ecotypes is important from an ecological point of view. If we look at slow-growing ubiquitous prokaryotes like *Pelagibacter ubique* and the marine *Crenarchaea* group 1, the differences between clones in the same sample seem to be much lower than in the case of fast-growing r-strategists like *S. baltica* (Ziemke et al. 1997, Garcia-Martinez & Rodriguez-Valera 2000). To determine whether this is a general difference between k- and r-strategists, further investigations of the fine structure of bacterioplankton species composition over time and space are needed. Future comparative genomics might provide more insight into this question by teaching us how genome-shaping forces influence intraspecies diversity or, vice versa, how intraspecies diversity influences genome organization.

Dynamics of bacterial communities can provide another insight into the question of what taxonomic level is needed to link structure and function, provided that these dynamics are followed at a high taxonomic resolution (e.g. by high resolution fingerprints or by clone libraries). Fuhrman et al. (2006) provided new insight into bacterial community dynamics, including verified predictions, using high resolution fingerprinting. In their study, patterns in bacterial community dynamics re-occurred at the species level along with environmental conditions over the course of 4 yr. For a specific biogeochemical function, an individual species, most closely related to *Thiomicrospira denitrificans*, was shown to drive autotrophic denitrification in the central Baltic (Brettar et al. 2006). These findings, based on the studies of natural communities, also support the notion that molecular studies on bacterioplankton diversity should reach the species level in terms of taxonomic resolution.

The other side of the key problem of linking microbial community structure with biogeochemical function is the precise definition of a single biogeochemical process. One of the most well-understood examples may be oxygenic photosynthesis by *Cyanobacteria*. In the case of *Prochlorococcus*, we even know at the subspecies level that various genetically well-defined ecotypes occupy different niches in the water column and in specific biomes of the world ocean (Zwirgmaier et al. 2008). If we look at the heterotrophic side of the aquatic carbon cycle, the situation is almost infinitely more complex, i.e. thousands of bacterial species can use dissolved organic carbon (DOC) and mineralize it to CO₂. On the other hand, there is strong evidence that, in open ocean habitats, *Pelagibacter ubique* catalyzes a large part of DOC flux to CO₂ (Malmström et al. 2005). Recent genomic data suggest that this carbon flux, like oxygenic photosynthesis by *Prochlorococcus*, is catalyzed to a major extent by only a few abundant ecotypes of *P. ubique* occurring in different oceanic provinces (Wilhelm et al. 2007).

LINKING BIOGEOCHEMICAL FUNCTION AND MICROBIAL DIVERSITY

Several approaches, usually combining molecular and chemical techniques, have been developed to relate biogeochemical function to specific types of bacteria. We evaluate the utility and the potential of these approaches using the following criteria: (1) the phylogenetic resolution of the technique, (2) the potential to study the process under close to *in situ* conditions, (3) the potential to monitor the process over time, and (4) the potential to monitor the overall bacterial community. The close to *in situ* conditions are considered to be very crucial for identifying key catalysts. Only under *in situ* conditions can it be expected that the 'right' bacteria perform the 'right' process. *In situ* conditions also include the whole bacterial community, because interactions and competition effects among bacteria have to be considered as influencing the process and the activity of the respective catalyst. Therefore, monitoring the overall bacterial community during the study of a biogeochemical process is important. An overview of this evaluation is given in Table 1.

Several very powerful technologies developed in the last 10 yr, which try to link structure to function, are now succeeding because of new sequencing technologies. An overview of the potential and technological details of SIP is given by Neufeld et al. (2007) and Wagner et al. (2006). In SIP approaches, ¹³C or ¹⁵N-labeled compounds are used to label either DNA, RNA or phospholipid acids. The labeled DNA and RNA is then separated from the non-labeled nucleic acids and analyzed for the phylogenetic structure of the labeled bacteria. The phospholipids are analyzed directly—without any separation step—by gas chromatography-mass spectrometry (GC-MS), providing a higher sensitivity than the RNA and DNA approaches. However, this approach suffers from the constraint that phospholipids have a limited phylogenetic resolution. Using SIP technology in oligotrophic pelagic environments is considered to be difficult, because the needed substrate concentrations and incubation times may change the *in situ* processes and bacterial community too much.

All probe-based methods at the single cell level (e.g. FISH), as well as macro- or microarray-based methods, are excellent tools for studying a process when the taxonomic position of the bacterial catalysts is known. Additionally, they can provide a first insight into the phylogenetic environment of the bacterial catalysts when using an adequate set of probes. One example is the combination of FISH with microautoradiography, MAR-FISH (Lee et al. 1999) or Micro-FISH (Cottrell & Kirchman 2000). The most important advantage of Micro-FISH is that it provides insight into the incor-

Table 1. Summary of techniques providing information on structure–function relationships in aquatic microbial communities, including their phylogenetic resolution and potential limitations. SIP: stable isotope probing, MAR-FISH: microautoradiography (MAR) coupled with fluorescence *in situ* hybridization, MIMS: multi-isotope Imaging mass spectrometry, MICRO-CARD-FISH: FISH combined with catalysed reporter deposition (CARD) and MAR. Habitat annotated with P: pelagic and S: sediment

Technology	Description	Phylogenetic resolution	Limitations and advantages	Habitat	Close to <i>in situ</i> conditions?	Source
SIP	Addition of stable isotope compound (^{13}C , ^{15}N) to bacterial community					
DNA-SIP	Separation of marked DNA by density centrifugation	Dependent on NA analysis \Rightarrow all resolution feasible	Long incubation times, requires addition of amounts of substrate	S	No (enrichment)	Neufeld et al. (2007)
RNA-SIP	Separation of marked RNA by density centrifugation	Dependent on NA analysis \Rightarrow all resolution feasible	Less incubation time and less substrate addition than SIP	S	Yes	Manefield et al. (2002)
PLFA-SIP	Mass spec analysis of PLFA	Dependent on database \Rightarrow in general, more difficult than DNA based	Sensitive SIP method, but less versatile with respect to phylogenetic analysis	S	Yes	Boschker & Middelburg (2002)
MAR-FISH	Laser confocal microscopy of samples after incubation with radioactively labelled substrate, use of fluorescent oligonucleotide probe	Probe-dependent		S	Yes	Wagner (2004)
MICRO-CARD-FISH +	Laser confocal microscopy of samples after incubation with radioactively labelled substrate, use of fluorescent oligonucleotide probe	Probe-dependent	Highly sensitive and accurate method, but less versatile with respect to phylogenetic analysis	P, S	Yes	Sintes & Herndl (2006)
Isotope array	Microarray analysis of samples after incubation with radioactively labelled substrate	Probe-dependent		S	Yes	Wagner et al. (2006)

Table 1 (continued)

Technology	Description	Phylogenetic resolution	Limitations and advantages	Habitat	Close to <i>in situ</i> conditions?	Source
Finger-print-based approaches	Stimulation experiments + fingerprint, fingerprint-based primer design followed by qPCR, FISH and generation of full 16S rRNA sequence	Any resolution feasible		P, S	Yes	Brettar et al. (2006)
MIMS	Secondary ion mass spec analysis of single cells (beam size of 33 nm, 1–3 atomic layers); addition of stable isotope substrate or study of natural isotope abundance	Probe-dependent (can be isotopically labelled)	Highly sensitive method (1000× that of ¹⁴ C-autoradiography) Less versatile with respect to phylogenetic analysis Very expensive!	P, S	Yes	Lechene et al. (2006), McMahon et al. (2006)

poration of radiolabeled compounds at the single cell level, i.e. on cells that can be labeled with ³H, ¹⁴C, ³⁵S or ³³P. Even the possibility of quantitative activity assessment at the single cell level has been demonstrated using CARD-FISH (Sintes & Herndl 2006).

A very promising new tool, which is analogous to microautoradiography, is multi-isotope imaging mass spectrometry (MIMS). MIMS is a new generation of secondary ion mass spectrometry (SIMS) with sophisticated ion optics and quantitative image analysis software (Kuypers & Jørgensen 2007). After incubation with stable isotope-containing substrates, the abundance of the isotope in cells that incorporated the substrates is analyzed at a very high spatial (33 nm, depth resolution: <1 nm) and mass resolution, with very high sensitivity and reproducibility. The possibilities of this new tool have been extensively demonstrated by Lechene et al. (2006) and McMahon et al. (2006). For example, MIMS was used to examine ¹⁵N₂ uptake by nitrogen-fixing bacteria at the subcellular level (Lechene et al. 2007).

Fingerprint-based approaches have the advantage of providing an overview of the bacterial community while the biogeochemical process proceeds. To analyze the bacterial catalysts of autotrophic denitrification and sulfur oxidation in the pelagic zone of the central Baltic Sea, Brettar et al. (2006) used a fingerprint-based approach to analyze samples from stimulation experiments. After autotrophic denitrification was stimulated by the addition of thiosulfate, this biogeochemical process was followed by measuring the uptake of radiolabeled bicarbonate as well as nitrate consumption. Concomitantly, bacterial community dynamics were analyzed by fingerprints (Fig. 8). The fingerprints showed the bacterial catalyst as a highlighted band, while the overall bacterial community remained stable. Sequencing of the highlighted band provided the basis for the construction of highly specific primers that were used to obtain the full 16S rRNA gene sequence of the bacterial catalyst (Höfle et al. 2005), and to quantify the bacterial catalyst abundance (DNA) by real-time PCR and assess its activity (rRNA abundance, assessed by real-time RT-PCR) *in situ* and under experimental conditions (Labrenz et al. 2004). Additionally, the sequence data were used to design probes for Micro-FISH analyses.

In general, biogeochemical stimulation experiments can provide a good basis for the study of the process of interest. Stimulation experiments are designed so that the biogeochemical process

of interest is enhanced, but the overall *in situ* conditions, including the bacterial community, are maintained. Many of the methods listed above can be used to analyze stimulation experiments in order to identify the bacteria catalyzing the specific process. The most challenging prerequisite for stimulation experiments is a good knowledge of the process and its kinetics as well as the required *in situ* conditions — and the possibility to mimic or to maintain them. Major efforts are still needed concerning the study of biogeochemical processes as well as in developing adequate technologies in order to mimic or maintain *in situ* conditions.

DRIVERS AND REGULATORY FACTORS OF BACTERIOPLANKTON DIVERSITY AND BIOGEOCHEMICAL FUNCTIONING

Bacterioplankton is subjected to 3 major drivers according to current understanding in aquatic ecology: (1) nutrients, (2) grazers, and (3) viruses. Since these, and other drivers like temperature, light, etc., have been studied from many aspects and an enormous wealth of knowledge is available, we concentrate on the relationship of the 3 major drivers of bacterioplankton community structure and how they influence its biogeochemical functioning. For other aspects, we refer to recent reviews (Pernthaler 2005, Karl 2007, Suttle 2007).

For bacterioplankton, nutrients are either organic or inorganic and either dissolved (dissolved organic matter, DOM) or particulate (particulate organic matter, POM). The time scales of nutrient dynamics relevant to the bacterioplankton community range from short pulses of nutrients over minutes or hours to long-term changes due to eutrophication or climate change with variations over years and decades. Organic nutrient pulses can be simulated well in experiments, and one of the first studies of bacterioplankton community dynamics indicated that the addition of readily available organic carbon to lake water mesocosms resulted in a strong increase of *Aeromonas hydrophila* (Höfle 1992). This observation was later confirmed by several studies and generalized in the 'gamma-shift' hypothesis, stating that the addition of DOC to aquatic bacterial communities induces a community shift towards the rapid responders, which are in general mostly members of the *Gammaproteobacteria* such as *Alteromonas*, *Shewanella*, etc. (Fuchs et al. 2000). In more general terms, this decrease in bacterioplankton richness under more eutrophic conditions was demonstrated by Reinthaler et al. (2005). In that study, richness of bacterioplankton in the North Sea was assessed by terminal restriction fragment length polymorphism (T-RFLP), and bacterial production and respiration were determined over a full

seasonal cycle. The lowest number of terminal restriction fragments (T-RFs) was observed at the highest bacterial production and respiration values. In addition, high bacterial growth efficiency was coupled with the lowest number of T-RFs, and vice versa. How universal these findings are will have to be examined in other ecosystems, because other factors, such as temperature and light intensity, might have played a significant role in the North Sea and cannot be excluded as overwriting factors for the described effects.

Next to bottom-up effects such as pulses of organic nutrients, top-down effects by grazers strongly influence the bacterioplankton community structure (Hahn & Höfle 2001, Pernthaler 2005). Grazing has a well-known effect on the morphological structure of bacterioplankton, i.e. it reduces the medium-size bacterioplankton and promotes the occurrence of ultra-microbacteria and very large bacterial cells that are resistant to grazing. Grazing can generate blooms of specific grazing-resistant bacterial species that can contribute up to 40% of the total biomass of bacterioplankton communities (Pernthaler et al. 2004). On the other hand, nutrient pulses can induce increased abundances of rapidly growing bacteria that are immediately grazed down by heterotrophic nanoflagellates (HNF), as was the case for *Aeromonas hydrophila* reported above (Höfle 1999). The selectivity of grazing also depends on the type of grazer; for example, daphnids are known to graze almost all bacterial species down to a certain size, whereas HNFs can graze certain bacterial species they prefer selectively (Hahn & Höfle 2001). In addition, a multitude of bacterial species have a variety of defense mechanisms against grazing, ranging from increased motility to the formation of biofilms (Matz & Kjelleberg 2005). All in all, the interactions of grazers with bacterioplankton depend on the taxonomic structure of both communities, i.e. the predator community as well as the prey community.

Viruses, the second important biological driver affecting bacterioplankton, can have a strong impact on the community composition and, due to lysis, on the biogeochemical cycling of organic matter via the viral shunt (Suttle 2007). In addition to these 2 functions comparable to the influence of grazers, viruses have a substantial effect on the genetic evolution of bacterioplankton as a prime agent of horizontal gene transfer by transduction (Weinbauer & Rassoulzadegan 2004, Medini et al. 2005). Here, we concentrate on the impact of viruses on the community structure of bacterioplankton and refer to the comprehensive review by Suttle (2007) for other functions of aquatic viruses. The general understanding of virus–bacteria interactions in planktonic communities is based on the 'killing the winner' model, in which the most abundant members

of the bacterioplankton community are infected and lysed by specific phages (Thingstad 2000). However, many cases demonstrate that blooms of bacteria persist despite high titers of viruses, or that less abundant bacteria are controlled by viruses (Waterbury & Valois 1993, Bouvier & del Giorgio 2007). Crucial for the understanding of these virus–bacteria interactions is, similar to the interaction with grazers, the composition of both communities. Understanding of the molecular diversity of viroplankton has progressed in a way similar to that of bacterioplankton, with the recent metagenomic analyses pointing at an even larger diversity among different water samples than that found for bacterioplankton (Angly et al. 2006, Culley et al. 2006). Matching studies using high resolution methods for the assessment of molecular diversity of both communities are still lacking, but would help to understand the virus–bacteria interactions in detail. All these interactions might depend on transient matches between both communities, and it would therefore help to study these interactions along large scales over time and space.

Looking at bacterioplankton communities on a global scale, the 4 major ocean domains or biomes (the Coastal Boundary Domain, Polar Domain, Westerly Winds Domain and Trade Winds Domain) as defined by Longhurst (1995) have to be taken into account. It is clear that very different environmental conditions occur in these 4 biomes and, therefore, different bacterioplankton communities are encountered, hinting on the regulatory factors of the structure and composition of these communities. An example for such a global comparison is the recent study by Pommier et al. (2007) comparing 9 bacterioplankton communities from all 4 oceanic biomes using 16S rRNA clone libraries. This study suggested that most of the rare members of the community were endemic and several of the abundant members were globally distributed. How closely related the global members are at the ecotype level is not clear from this study, because only around 500 bp of the 16S rRNA gene sequence were examined and it lacks resolution in terms of abundance assessment as explained above. The global distribution and abundance of ecotypes of *Synechococcus* and *Prochlorococcus* were assessed in a recent study by Zwirgmaier et al. (2008). This study is an excellent example in demonstrating how various ecotypes of both cyanobacteria correspond with comparable environmental conditions and the occurrence of different signature ecotypes or ecotype consortia in the 4 major ocean biomes.

How much physical resolution is needed to assess the global pattern of bacterioplankton communities might depend on the physico-chemical conditions found in specific parts of the world ocean. For example, the central part of the Baltic Sea is characterized by a

bacterioplankton community that is rather similar across several hundred kilometres on the surface (M. G. Höfle pers. obs.), but the community structure and composition change with depth within a few metres at the oxic–anoxic interface around 130 m (Brettar et al. 2006). In addition to the spatial extent, the time scale is an important aspect. Annual cycles are of high relevance in temperate regions and are reflected by substantial changes in bacterioplankton community structure (Höfle et al. 1999); however, even in moderate, coastal environments reoccurring seasonal patterns were observed and helped to reveal factors regulating the bacterioplankton community composition (Fuhrman et al. 2006). Long-term analyses over many years and decades are very relevant for the understanding of eutrophication and climate change, but are quite challenging to conduct due to the lack of planktonic sample material spanning larger time spans.

We look at diversity from 2 perspectives — the bacteria and ecosystem functioning. For the bacterial perspective, a crucial point is intraspecies diversity. First, we have to differentiate between intraspecies diversity within the same water mass, i.e. competition and interaction of populations of the same species, and second, between intraspecies diversity at different times and in different water masses (space), i.e. without potential competition and interaction, that can be regarded as adaptation to the present conditions. Though there is competition for the same resources, intraspecies diversity may provide advantages such as higher protection for the species, e.g. by the presence of populations within a species with different viral resistance, allowing for survival of some populations under virus attack, or the use of different substrates by a population while other substrates are more efficiently consumed by competing species. In addition, collaborative use of substrates among different populations might be advantageous. In general, the provision of a high genetic recombination potential can be advantageous by enabling the species to adapt to a rapidly changing environment. A large pan genome can thus be a great evolutionary advantage for a species. In contrast, high intraspecies diversity seems to be less needed in a comparatively constant environment.

From an ecosystem functioning perspective, intraspecies and interspecies diversity may fulfill similar tasks, i.e. lead to an efficient energy and matter flow and turnover due to the shaping forces of intraspecies and interspecies interaction and competition. Resilience of bacterial populations can be considered as an important mechanism underlying stable biogeochemical cycling. How much intraspecies and interspecies interactions are contributing to the resilience of the whole bacterial community over time and space is still not clear and remains an interesting question. In

general, we assume that biogeochemical cycling is greatly stabilized by intraspecies and interspecies interactions and the overall genetic diversity of the whole microbial community.

CONCLUSIONS AND FUTURE PERSPECTIVES

Community structure and composition of bacterioplankton

Bacterioplankton community structure is characterized by the 'abundant & rare member' model of its rank-abundance distribution according to current molecular evidence (Pedrós-Alió 2006, Huber et al. 2007). This model means that, if we look at an average 1 l sample of natural surface water containing 10^9 prokaryotic cells, more than half of the cells ($>5 \times 10^8$) would comprise only a few abundant species. The second half of the sample might consist of 10s to 100s of less abundant species. The last milliliter with its 10^6 cells could contain 100 000 species, each represented by only 10 cells (see bottom lines in Fig. 3). This seedbank could be the 'everything' in Baas-Beeking's (1934) concept 'everything is everywhere—the environment selects' (Suttle 2007). This understanding of bacterioplankton community structure and the reoccurrence of specific community compositions points to the following directions for future research: (1) identification, isolation and physiological characterization of the abundant members of the community, (2) assessment of structure and composition of the seedbank, (3) identification of the mechanism triggering the transition from rare to abundant member and vice versa, and (4) understanding of interactions of the bacterioplankton community with the whole aquatic food web and its physico-chemical environment. All these aspects of diversity should be examined in the light of its biogeochemical relevance as detailed in 'Predicting biogeochemical functioning of pelagic ecosystems' below.

Identification, isolation and genome analysis of the abundant species, in combination with determination of the environmental conditions, provides information of the genetic make-up of a core community and how it reflects the environmental conditions. The abundant species are most likely the key catalysts for major biogeochemical processes in the water column. Isolates of abundant members provide information on genome sequences, clonal structure and physiology of these species. The success of such an approach was very recently demonstrated by discovering a fifth biochemical pathway to fix CO_2 in *Archaea* and understanding the function and high abundance of the respective genes in metagenome data from the GOS database (Berg et al. 2007). More information on the rare mem-

bers, to be expected from new sequencing approaches and improved bioinformatics, could help to understand the whole genetic potential of the bacterioplankton community. Understanding of the transition mechanisms from rare to abundant members would provide insights into how this genetic potential is regulated and turned into the observed biogeochemical activities. Understanding the interactions with the whole aquatic ecosystem on a global scale and at appropriate time intervals is important. Global sample archives, long-term research sites and an international effort are needed to understand the world-wide pattern and dynamics of bacterioplankton communities (Karl 2007). Another issue deserving substantial efforts is the comparison of methods. Currently, no datasets exist that compare pyrosequencing, PCR cloning and fingerprints with non-PCR based metagenomic analyses. Detailed analyses of the pitfalls and biases of all these molecular methods cannot be properly performed without such a comparative identification of possible problems or artifacts. The same is true for 'universal' or specific PCR primers or probes. While almost all our studies on microbial diversity depend on the accuracy of this approach, there are no dedicated studies combining bioinformatics analyses and experiments to verify how effective these primers are in binding and amplifying different sequences.

Predicting biogeochemical functioning of pelagic ecosystems

When the concept of a genome-enabled biogeochemistry was put forward several years ago, it was based on very simple microbial communities such as biofilms of acid mine drainage (Newman & Banfield 2002, Tyson et al. 2004). Extending this concept to pelagic ecosystems and their functional understanding means adding several levels of complexity (Falkowski et al. 2008). Therefore, we categorize the structure–function relationships into what we think are the 3 most relevant cases: (1) single biogeochemical processes with multiple microorganisms, (2) multiple biogeochemical processes with single microorganisms, and (3) multiple biogeochemical processes with multiple microorganisms.

Ammonium oxidation provides a good example for a single process that is catalyzed by several groups of prokaryotes, i.e. *Crenarchaea*, *Betaproteobacteria* and *Gammaproteobacteria*. The major biochemical reactions in this process are the same, but there are major differences among the key enzymes catalyzing ammonium oxidation (Könneke et al. 2005, Wuchter et al. 2006). In addition, the whole metabolic network in which these key enzymes are integrated are different,

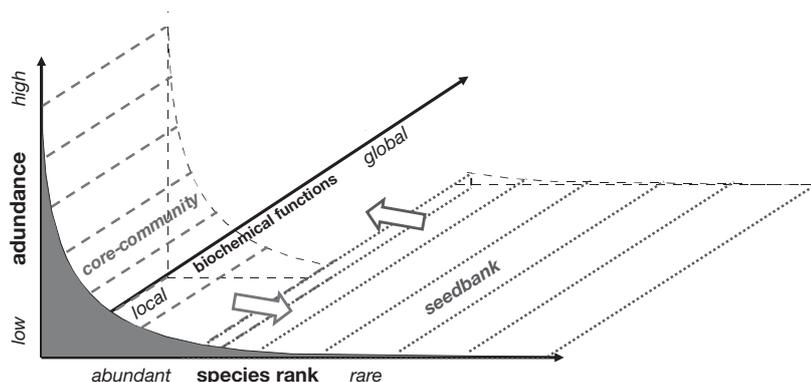


Fig. 9. Conceptual model of pelagic microbial communities and their biogeochemical functioning in aquatic ecosystems

especially if we compare archaeal with bacterial ammonium oxidation (Könneke et al. 2005, Berg et al. 2007). Therefore, it has to be expected that different key catalysts, i.e. species of prokaryotes, dominate ammonium oxidation in the water column depending on the prevailing physico-chemical conditions (Francis et al. 2005, Caffrey et al. 2007).

The examples of *Pelagibacter ubique* and *Aeromonas hydrophila* demonstrate the case of multiple biogeochemical processes, such as mineralization of complex DOC, which are dominated by a single bacterial species. These examples illustrate especially the dynamics of any of these biogeochemical functions. In the case of *P. ubique*, it can be assumed that the species evolved over eons of time to dominate the open sea as a k-strategist, whereas many pelagic r-strategists such as *A. hydrophila* or *Shewanella baltica* may only occur in small abundances in the rare fraction of the community until a phytoplankton bloom supplies organic nutrients. The third case might be the most 'common' in bacterioplankton communities, i.e. many different bacterial species consume the DOC, as recently demonstrated by Mou et al. (2008).

In more general terms, one could assume that the major biogeochemical functions and fluxes in a water sample of the pelagic zone are performed by the abundant members of the bacterioplankton, which we call core community. Over time and space, these functions and fluxes change and some of the rare members might replace some of the abundant members and become abundant themselves. Considering evolutionary time frames, the huge population sizes of individual pelagic species, and the interconnectivity of the world ocean, we hypothesize that the seedbank might contain a certain share of ubiquists. How large this share is and who they are is currently one of the most interesting questions. The whole prokaryotic community in a water sample, i.e. the core community plus the seedbank, could be called pan community, in analogy

to the pan genome of a species. This pan community should be able to perform all the biogeochemical functions that will occur in pelagic systems over time and space in an optimal manner. A schematic view of the whole concept of core and pan genome and its relationship to the biogeochemical functioning of bacterioplankton in the water column is given in Fig. 9.

Today, we can look back at about 20 yr of molecular analyses of bacterioplankton and realize that we have made some progress and developed an enormous array of techniques. Some of the most sophisticated techniques, like pyrosequencing and MIMS, are very promising with respect to answering many questions about microbial communities and their biogeochemical functioning. From an ecological perspective, we think that now is the time to take advantage of the technological progress in a holistic way, because the biogeochemical functioning is an emerging property of the whole bacterial community and not of its individual members (Laughlin 2005, Follows et al. 2007). This holistic perspective might provoke the following future trends in bacterioplankton research: (1) global microbial inventories and across-ecosystem comparisons, (2) comparative metagenomics and systems biology of major communities, (3) combined studies of biogeochemical and molecular analyses, and (4) integration of biogeochemical processes at the micro-, meso- and global scale. These trends might lead to the development of a community theory for bacterioplankton that could explain its global biogeochemical functioning.

Acknowledgements. We thank G. Herndl very much for critically reviewing the manuscript and stimulating discussion. The help by K. Henne with the graphical presentations is greatly acknowledged. This work was supported by funds to M.G.H. from the European Commission for the HEALTHY WATER project (FOOD-CT-2006-036306). The authors are solely responsible for the content of this publication. It does not represent the opinion of the European Commission. The European Commission is not responsible for any use that might be made of data appearing herein.

LITERATURE CITED

- Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143–169
- Angly FL, Felts B, Breitbart M, Salamon P and others (2006) The marine viromes of 4 oceanic regions. *PLoS Biol* 4:e368
- Baas-Becking LGM (1934) *Geobiologie of inleiding tot de milieukunde*. WP Van Stockum & Zoon, The Hague
- Béjà O, Aravind L, Koonin EV, Suzuki MT and others (2000)

- Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289:1902–1906
- Berg IA, Kockelkorn D, Buckel W, Fuchs G (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in *Archaea*. *Science* 318:1782–1785
- Boschker HTS, Middelburg JJ (2002) Stable isotopes and biomarkers in microbial ecology. *FEMS Microbiol Ecol* 40:85–95
- Bouvier T, del Giorgio PA (2007) Key role of selective viral-induced mortality in determining marine bacterial community composition. *Environ Microbiol* 9:287–297
- Brettar I, Höfle MG (1992) Influence of ecosystematic factors on survival of *Escherichia coli* after large-scale release into lake water mesocosms. *Appl Environ Microbiol* 58:2201–2210
- Brettar I, Christen R, Höfle MG (2002a) *Rheinheimera baltica* gen. nov., sp. nov., a blue coloured bacterium isolated from the central Baltic sea. *Int J Syst Evol Microbiol* 52:1851–1857
- Brettar I, Christen R, Höfle MG (2002b) *Shewanella denitrificans*, sp. nov., a vigorously denitrifying bacterium isolated from the oxic-anoxic interface of the Gotland Deep in the central Baltic Sea. *Int J Syst Evol Microbiol* 52:2211–2217
- Brettar I, Labrenz M, Flavier S, Bötterl J, Kuosa H, Christen R, Höfle MG (2006) Identification of a *Thiomicrospira denitrificans*-like epsilonbacterium as a catalyst for denitrification in the central Baltic sea. *Appl Environ Microbiol* 72:1364–1372
- Caffrey JM, Bano N, Kalanetra K, Hollibaugh JT (2007) Ammonia oxidation and ammonia-oxidizing bacteria and archaea from estuaries with differing histories of hypoxia. *ISME J* 1:660–662
- Campbell BJ, Waidner LA, Cottrell MT, Kirchman DL (2008) Abundant proteorhodopsin genes in the North Atlantic Ocean. *Environ Microbiol* 10:99–109
- Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB, Welschmeyer NA (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* 334:340–343
- Coenye T, Gevers D, Van de Peer Y, Vandamme P, Swings J (2005) Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* 29:147–167
- Cohan FM (2002) What are bacterial species? *Annu Rev Microbiol* 56:457–487
- Cottrell MT, Kirchman DL (2000) Natural assemblages of marine proteobacteria and *Cytophaga-Flavobacter* consuming low and high-molecular weight dissolved organic matter. *Appl Environ Microbiol* 66:1692–1697
- Cottrell MT, Moore JA, Kirchman DL (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* 65:2553–2557
- Cottrell MT, Waidner LA, Yu L, Kirchman DL (2005a) Bacterial diversity of metagenomic and PCR libraries from the Delaware River. *Environ Microbiol* 7:1883–1895
- Cottrell MT, Yu L, Kirchman DL (2005b) Sequence and expression analyses of *Cytophaga*-like hydrolases in a western Arctic metagenomic library and the Sargasso Sea. *Appl Environ Microbiol* 71:8506–8513
- Culley AI, Lang AS, Suttle CA (2006) Metagenomic analysis of coastal virus communities. *Science* 312:1795–1798
- DeLong EF (2005) Microbial community genomics in the ocean. *Nat Rev Microbiol* 3:459–469
- Deng J, Auchtung JM, Rodrigues JLM, Tiedje JM (2008) *Shewanella baltica*: a model for examining specialization along a redox gradient. Joint Genomics: GTL Contractor-Grantee Workshop VI & Metabolic Engineering Working Group Interagency, Conference on Metabolic Engineering 2008, Bethesda, MD
- Dufresne A, Salanoubat M, Partensky F, Artiguenave F and others (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxygenic phototrophic genome. *Proc Natl Acad Sci USA* 100:10020–10025
- Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M and others (2006) Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 7:57–70
- Falkowski PG, Fenchel T, DeLong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* 320:1034–1039
- Follows MJ, Dutkiewicz S, Grant S, Chisholm SW (2007) Emergent biogeography of microbial communities in a model ocean. *Science* 315:1843–1846
- Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB (2005) Ubiquity and diversity of ammonia-oxidizing *Archaea* in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* 102:14683–14688
- Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, DeLong EF (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* 105:3805–3810
- Fuchs BM, Zubcov MV, Sahm K, Burkil PH, Amann RI (2000) Changes in community composition during dilution cultures of marine bacterioplankton as assessed by flow cytometric and molecular biological techniques. *Environ Microbiol* 2:191–201
- Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* 103:13104–13109
- Garcia-Martinez J, Rodriguez-Valera F (2000) Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine *Archaea* of group I. *Mol Ecol* 9:935–946
- Gevers D, Cohan FM, Lawrence JG, Spratt BG and others (2005) Re-evaluating prokaryotic species. *Nat Rev Microbiol* 3:733–739
- Giovannoni S, Stingl U (2007) The importance of culturing bacterioplankton in the 'omics' age. *Nat Rev Microbiol* 5:820–826
- Giovannoni SJ, Tripp HJ, Givan S, Podar M and others (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309:1242–1245
- Hahn MW, Höfle MG (2001) Grazing of protozoa and its effect on populations of aquatic bacteria. *FEMS Microbiol Ecol* 35:113–121
- Hahn MW, Lünsdorf H, Lu Q, Schauer M, Höfle MG, Boenigk P, Stadler P (2003) Isolation of novel ultramicrobacteria belonging to the *Actinobacteria* from five freshwater habitats in Europe and Asia. *Appl Environ Microbiol* 69:1442–1451
- Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* 210:1518–1525
- Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68:669–685
- He Z, Gentry TJ, Schadt CW, Wu L and others (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* 1:67
- Hirsch P (1979) Life under conditions of low nutrient concentrations (group report). In: Shilo M (ed) *Strategies of microbial life in extreme environments*. Verlag Chemie, Weinheim, p 357–372
- Höfle MG (1992) Bacterioplankton community structure and

- dynamics after large-scale release of nonindigenous bacteria as revealed by low-molecular-weight RNA analysis. *Appl Environ Microbiol* 58:3387–3394
- Höfle MG (1999) Direct detection of nutrient induced changes in the community structure of bacterioplankton using low-molecular-weight RNA analysis. *Arch Hydrobiol* 54: 199–211
- Höfle MG, Haas H, Dominik K (1999) Seasonal dynamics of bacterioplankton community structure in a eutrophic lake as determined by 5S rRNA analysis. *Appl Environ Microbiol* 65:3164–3174
- Höfle MG, Ziemke F, Brettar I (2000) Niche differentiation of *Shewanella putrefaciens* populations from the Baltic as revealed by molecular and metabolic fingerprinting. In: Bell CR, Brylinsky M, Johnson-Green P (eds) *Microbial biosystems: new frontiers*. Proceedings of ISME8, Atlantic Canada Society for Microbial Ecology, Halifax, Canada, p 135–143
- Höfle MG, Flavier S, Christen R, Bötzel J, Labrenz M, Brettar I (2005) Retrieval of nearly complete 16S rRNA gene sequences from environmental DNA following 16S rRNA-based community fingerprinting. *Environ Microbiol* 7: 670–675
- Huber JA, Welch DBM, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML (2007) Microbial population structures in the deep marine biosphere. *Science* 318: 97–100
- Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Res* 17:377–386
- Karl DM (2007) Microbial oceanography: paradigms, processes and promise. *Nat Rev Microbiol* 5:759–769
- Könneke M, Bernhard AE, de la Torre JR, Walker CM, Waterbury JB, Stahl DA (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437: 543–546
- Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 102:2567–2572
- Kuypers MMM, Jørgensen BB (2007) The future of single-cell environmental microbiology. *Environ Microbiol* 9:6–7
- Labrenz M, Brettar I, Flavier S, Christen R, Bötzel J, Höfle MG (2004) Development and application of a real-time PCR approach for quantification of uncultured bacteria in the central Baltic Sea. *Appl Environ Microbiol* 70:4971–4979
- Laughlin RE (2005) A different universe—reinventing physics from the bottom down. Basic Books, New York
- Lechene C, Hillion F, McMahon G, Benson D and others (2006) High-resolution quantitative imaging of mammalian and bacterial cells using stable isotope mass spectrometry. *J Biol* 5:20
- Lechene CP, Luyten Y, McMahon G, Distel DL (2007) Quantitative imaging of nitrogen fixation by individual bacteria within animal cells. *Science* 317:1563–1566
- Lee N, Nielsen PH, Andreassen KH, Juretschko S, Nielsen JL, Schleifer KH, Wagner M (1999) Combination of fluorescent *in situ* hybridization and microautoradiography—a new tool for structure-function analyses in microbial ecology. *Appl Environ Microbiol* 65:1289–1297
- Lindstedt BA (2005) Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* 26:2567–2582
- Liu Z, Lozupone C, Harmady M, Bushman FD, Knight R (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* 35:e120
- Longhurst A (1995) Seasonal cycles of pelagic production and consumption. *Prog Oceanogr* 36:77–167
- Malmström RR, Cottrell MT, Elifantz H, Kirchman DL (2005) Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Appl Environ Microbiol* 71:2979–2986
- Malmström RR, Straza TRA, Cottrell MT, Kirchman DL (2007) Diversity, abundance, and biomass production of bacterial groups in the western Arctic Ocean. *Aquat Microb Ecol* 47:45–55
- Manefield M, Andrew S, Whiteley R, Griffiths RI, Bailey MJ (2002) RNA stable isotope robing, a novel means of linking microbial community function to Phylogeny. *Appl Environ Microbiol* 68:5367–5373
- Marcy Y, Ouverney C, Bik EM, Lösekann T and others (2007) Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci USA* 104:11889–11894
- Margulies M, Egholm M, Altman WE, Attiya S and others (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380
- Matz C, Kjelleberg S (2005) Off the hook—how bacteria survive protozoan grazing. *Trends Microbiol* 13:302–307
- McMahon G, Glassner BJ, Lechene CP (2006) Quantitative imaging of cells with multi-isotope imaging mass spectrometry (MIMS)—nanoautography with stable isotope tracers. *Appl Surf Sci* 252:6895–6906
- Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. *Curr Opin Genet Dev* 15:589–594
- Moran MA, Miller WL (2007) Resourceful heterotrophs make the most of light in the coastal ocean. *Nat Rev Microbiol* 5:792–800
- Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Keine RP (2004) Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* 432:910–913
- Mou X, Sun S, Edwards RA, Hodson RE, Moran MA (2008) Bacterial carbon processing by generalist species in the coastal ocean. *Nature* 451:708–711
- Neufeld JD, Wagner M, Murrell JC (2007) Who eats what, where and when? Isotope-labeling experiments are coming of age. *ISME J* 1:103–110
- Newman DK, Banfield JF (2002) Geomicrobiology: how molecular-scale interactions underpin biogeochemical systems. *Science* 296:1071–1077
- Oremland RS, Capone DG, Stolz JF, Fuhrman J (2005) Whither or wither geomicrobiology in the era of 'community metagenomics'. *Nat Rev Microbiol* 3:572–578
- Pedrés-Alió C (2006) Marine microbial diversity: can it be determined? *Trends Microbiol* 14:257–263
- Pernthaler J (2005) Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* 3:537–546
- Pernthaler J, Zöllner E, Warnecke F, Jürgens K (2004) Blooms of filamentous bacteria in a mesotrophic lake: identity and controlling mechanisms. *Appl Environ Microbiol* 70: 6272–6281
- Podar M, Abulencia CB, Walcher M, Hutchison D and others (2007) Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl Environ Microbiol* 73:3205–3214
- Pommier T, Canbäck B, Riemann L, Boström KH and others (2007) Global pattern of diversity and community structure in marine bacterioplankton. *Mol Ecol* 16:867–880
- Rappé MS, Connon SA, Vergin KL, Giovannoni SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418:630–633
- Reinthalder T, Winter C, Herndl GJ (2005) Relationship between bacterioplankton richness, respiration and pro-

- duction in the southern North Sea. *Appl Environ Microbiol* 71:2260–2266
- Roesch LFW, Fulthorpe RR, Riva R, Casella G and others (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* 1:283–290
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB and others (2007) The Sorcerer II Global Ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* 5:398–431
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* 74:5463–5467
- Schmidt TM (2006) The maturing of microbial ecology. *Int Microbiol* 9:217–223
- Schmidt TM, DeLong EF, Pace NR (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* 173:4371–4378
- Sintes E, Herndl GJ (2006) Quantifying substrate uptake by individual cells of marine bacterioplankton by catalyzed reporter deposition fluorescence *in situ* hybridization combined with microautoradiography. *Appl Environ Microbiol* 72:7022–7028
- Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* 8:732–740
- Sogin ML, Morrison HG, Huber JA, Welch DM and others (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* 103:12115–12120
- Stepanauskas R, Sieracki ME (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci USA* 104:9052–9057
- Sundquist A, Bigdeli S, Jalili R, Druzin ML and others (2007) Bacterial flora-typing with targeted, chip-based pyrosequencing. *BMC Microbiol* 7:108
- Suttle CA (2007) Marine viruses — major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812
- Thingstad TF (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
- Tyson GW, Chapman J, Hugenholtz P, Allen EE and others (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:37–43
- Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J (1996) Polyphasic taxonomy: a consensus approach to bacterial taxonomy. *Microbiol Rev* 60:407–438
- Venter JC, Remington K, Heidelberg JF, Halpern AL and others (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74
- von Wintzingerode F, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21:213–229
- Wagner M (2004) Deciphering functions of uncultured microorganisms. *ASM News* 2:63–70
- Wagner M, Nielsen PH, Loy A, Nielsen JL, Daims H (2006) Linking microbial community structure with function: fluorescence *in situ* hybridization-microautoradiography and isotope arrays. *Curr Opin Biotechnol* 17:83–91
- Waterbury JB, Valois FW (1993) Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* 59:3393–3399
- Weinbauer MG, Rassoulzadegan F (2004) Are viruses driving microbial diversification and diversity? *Environ Microbiol* 6:1–11
- Wilhelm LJ, Tripp HJ, Givan SA, Smith DP, Giovannoni SJ (2007) Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct* 2:27–40
- Wuchter C, Abbas B, Coolen MJL, Herfort L and others (2006) Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA* 103:12317–12322
- Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, Church GM (2006) Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* 24:680–686
- Ziemke F, Brettar I, Höfle MG (1997) Stability and diversity of the genetic structure of a *Shewanella putrefaciens* population in the water column of the central Baltic. *Aquat Microb Ecol* 13:63–74
- Ziemke F, Höfle MG, Lalucat J, Rosselo-Mora R (1998) Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica*, sp. nov. *Int J Syst Bacteriol* 48:179–186
- Zwirgmaier K, Jardillier L, Ostrowski M, Mazard S and others (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among ocean biomes. *Environ Microbiol* 10: 147–161

Submitted: February 12, 2008; Accepted: July 22, 2008

Proofs received from author(s): September 6, 2008