

# Biogeographic analysis of ribosomal RNA clusters from marine bacterioplankton

Thomas Pommier, Jarone Pinhassi, Åke Hagström\*

Marine Microbiology, Biology and Environmental Science, Kalmar University, 39182 Kalmar, Sweden

**ABSTRACT:** Bacterial cosmopolitanism, which has for a long time been accepted by most microbiologists, confronts recent reports of endemic species and restricted distributions. From a marine perspective, our current knowledge is lacking global pictures of bacterioplankton distribution. Public databases storing nucleotide sequences grow daily from recurrent environmental analyses of the microbial community using the small subunit ribosomal RNA gene (SSU rDNA) sequence as a marker of bacterial occurrence. These studies significantly contribute to our understanding of the global microbial processes. To address the question of whether marine bacterioplankton are cosmopolitan, 7070 SSU rDNA sequences submitted to GenBank were analyzed for sampling information and clustered into distinct ribotypes, based on sequence similarity levels. The resulting distribution maps displayed remarkable patterns of distributions of some members of the bacterioplankton community. Despite a strong bias toward sampling sites in the northern temperate regions of the globe, specific ribotypes showed ubiquitous dispersal whereas other ribotypes showed a distribution restricted to polar regions. These results suggest that marine bacterioplankton components present unexpected global patterns of distributions, where cosmopolitanism appears as an occasional trait.

**KEY WORDS:** Bacterioplankton · Global distribution · Cosmopolitanism · Ribotypes

*Resale or republication not permitted without written consent of the publisher*

## INTRODUCTION

During the last few decades, marine microbiologists focusing on biogeochemistry have demonstrated the importance of heterotrophic bacterioplankton in the cycling of matter (Azam 1998, Ducklow 2000). This has led to the need to identify the bacteria involved in the turnover of carbon and mineral nutrients. Because of this demand, and since a significant fraction of the bacterioplankton has failed to grow on conventional cultivation media, culture-independent techniques based on molecular biology have been developed. In particular, the small subunit ribosomal RNA gene (SSU rDNA) has been commonly used to link variations in the microbial community to observed changes in the environment (Acinas et al. 1999, Cottrell & Kirchman 2000, Hollibaugh et al. 2002). More recently, a few studies have provided elaborate pictures of the whole bacterioplankton community using large clone libraries of the SSU rDNAs (Acinas et al. 2004, O'Sullivan et al. 2004) or environmental genome shotgun sequencing (Venter

et al. 2004). After more than 2 decades, samples from almost everywhere can now be used to address questions on bacterioplankton global biogeography.

The common view that 'everything is everywhere', first formulated at the beginning of the last century by the biologist Beijerinck (Brock 1961), was newly re-introduced by Fenchel and Finlay in a series of publications (Finlay 1998, Finlay & Clarke 1999, Finlay 2002, Fenchel & Finlay 2004). Focused on the marine microbial community, support for the ubiquitous distribution of bacterioplankton came from group- or species-specific analyses. In the late 1970s (Waterbury et al. 1979), the widespread occurrence of unicellular marine planktonic bacteria from the genus *Synechococcus* was reported, and this has now been confirmed with advances in molecular biology (Carr & Mann 1994). By comparing SSU rDNA richness in distant coastal waters, several studies illustrated the concurrent presence of similar Archaea (DeLong 1992) and bacterioplankton phylotypes (Fuhrman et al. 1993, Mullins et al. 1995). Subsequently, similar isolates

\*Corresponding author. Email: ake.hagstrom@hik.se

collected around the globe presented similar SSU rDNAs (Hagström et al. 2000, Ward & O'Mullan 2002). The ubiquitous distribution of the SAR11 clade has emphasized its dominance within the marine bacterioplankton community (Morris et al. 2002). These reports of highly similar sequences from widely separated areas of the ocean have indicated that a few key microbial groups were widespread in their occurrence (Giovannoni & Rappé 2000).

Nevertheless, the century-old view of Beijerinck was recently challenged by several reports of allopatric speciation and opposing opinions on the ubiquity of small organisms. Coleman (2001) demonstrated a clear pattern of geographical localization of closely related protistan morphospecies of chlorophytes. Endemic strains of *Pseudomonas* were found in geographically distant soil samples from 4 continents (Cho & Tiedje 2000). One pattern of distribution was the bipolar occurrence of species in higher latitudes separated by a gap in distribution across the tropics such as that seen for identical genotypes of planktonic foraminifers (Darling et al. 2000). More recently, highly similar SSU rDNA sequences from ammonia-oxidizing bacteria were found both in the Arctic Ocean and in the Antarctic Palmer Peninsula (Hollibaugh et al. 2002), likely showing a bipolar distribution of this functional group. Diverse studies on marine bacterial species inhabiting extreme environments, e.g. hot springs (Papke et al. 2003) or sponge-associated bacteria (Taylor et al. 2005), show restricted distributions of some members of the free-living marine plankton community. Sea-ice bacteria (Staley & Gosink 1999) as well as archaeal assemblages (Massana et al. 2000) were shown to include diverging SSU rDNA variants from geographically separated areas, although some identical strains were also found. Related to its temperature optimum, *Prochlorococcus* has never been found at latitudes higher than 60°, thus showing a confined distribution to temperate and tropical regions (Partensky et al. 1999). The opposite was observed for a member of the *Roseobacter* clades, which was found to cluster in temperate and polar regions but was excluded from tropical regions (Selje et al. 2004).

Over the last 2 decades, the deposit of nucleotide sequences to public databases (e.g. GenBank) has created a large and valuable record of global marine bacterioplankton. In the present study, we investigated whether cosmopolitanism is a general pattern among free-living bacteria from the open-ocean and coastal communities of the global ocean. As in the study by Hagström et al. (2002), SSU rDNA sequences from all marine free-living bacterioplankton reported to GenBank were clustered into contigs to assess their global biogeography. As originally defined, a contig is 'a set of gel readings that are related to one another by

overlap of their sequences' (Staden 1979), i.e. a set of overlapping segments of DNA. Since in this study the focus was set on ribosomal DNA segments, each contig distinguishes a unique ribotype, including valuable metadata (e.g. sampling coordinates) to perform a distribution analysis (see Fig. 1). To organize the large dataset, we designed a local database, the Marine Bacterioplankton Database (MBD), to store all records from marine bacterioplankton SSU rDNAs and link them to their geographical occurrences. This database and the distribution maps of each cluster are publicly available at <http://halia.hik.se/mbd>.

## MATERIALS AND METHODS

**Data retrieving and contig assembly.** Free-living marine bacterioplankton SSU rDNA sequences and their GenBank-associated files were downloaded from the National Center for Biological Information website ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). The resulting dataset, which occasionally included sampling information, was imported into a local database (the MBD). Each sequence was checked against the search criteria and those not matching were removed. Sets of overlapping DNA sequences were assembled into contigs using the SeqManII software (Lasergene version 5, DNA Star). Fig. 1 presents an example of contig assembly. The retrieved records for this contig were complete or partial sequences of the SSU rDNA sequenced using different primer sets. In the present contig, the complete SSU rDNA was sequenced (sequence AF118019), and thus became the backbone of alignment of highly similar sequences. The assemblies were based on a  $\geq 100$  bp overlap and  $\geq 97$  or  $\geq 98\%$  similarity in overlapping regions. Even though the SeqManII software only allows a maximum of 50 bp overlap, each resulting contig was manually checked to ensure the overlap was over 100 bp. During the contig assembly, nucleotide insertions and gaps were limited to 70 per 1000 bases in both the aligned sequence and the whole contig. These standard settings constrained the assembly to an accurate alignment (i.e. long gaps were disabled). When they were not directly provided in the GenBank record, specific sampling coordinates were extracted from the corresponding references. Assignment of taxonomic identity for the consensus sequences was achieved by direct extraction from original data, i.e. when the sequences were submitted with their taxonomic affiliation. In the rare cases when different taxonomy was attributed to 2 sequences included in the same contig, which may be regarded as a chimera formation, the contig was de-assembled into 2 distinct ribotypes.

General information			Position in the contig (Numerotation in base pairs [bp])					Alignment information				Sampling information					
Accession	Length	Source	0	250	500	750	1000	1250	1500	Varying bases	Mismatch	Insertion	Gap	Coordinates	Depth (m)	Date	
AF118019	1497 bp	Vybiral	5' → 3'						6	11	1		Unknown	Unknown	Unknown		
AF025547	490 bp	Hagström	5' → 3'							2		1	43°41'N 07°19'E	0,5	Unknown		
X89346	227 bp	Benlloch	5' → 3'						2	5			43°41'N 03°54'E	Unknown	Unknown		
AF317677	992 bp	Leathers	3' ← 5'							2			Unknown	Unknown	Unknown		
AY095594	293 bp	Fuhrman	5' → 3'							5			32°56'N 120°42'W	3600	7 Apr 94		
AY095807	488 bp	Fuhrman	5' → 3'							1			16°35'S 166°25'E	4000	7 Apr 94		
AY095819	556 bp	Fuhrman	5' → 3'							2			16°35'S 166°25'E	4000	7 Apr 94		
AY095811	602 bp	Fuhrman	5' → 3'										16°35'S 166°25'E	4000	7 Apr 94		
AY095816	602 bp	Fuhrman	5' → 3'										16°35'S 166°25'E	4000	7 Apr 94		
AY095813	859bp	Fuhrman	5' → 3'							2		1	16°35'S 166°25'E	4000	7 Apr 94		
AY095794	853 bp	Fuhrman	5' → 3'										16°35'S 166°25'E	4000	7 Apr 94		
AY095799	854 bp	Fuhrman	5' → 3'							3			16°35'S 166°25'E	4000	7 Apr 94		
AY095801	854 bp	Fuhrman	5' → 3'							1			16°35'S 166°25'E	4000	7 Apr 94		
AY095818	849 bp	Fuhrman	5' → 3'							1			16°35'S 166°25'E	4000	7 Apr 94		
AY095806	579 bp	Fuhrman	5' → 3'										16°35'S 166°25'E	4000	7 Apr 94		
AY095815	579 bp	Fuhrman	5' → 3'										16°35'S 166°25'E	4000	7 Apr 94		
AY095906	579 bp	Fuhrman	5' → 3'										18°10'S 178°42'W	Unknown	8 Jun 91		
<i>Consensus 1499 bp</i>																	

— = consensus sequence is supported by only 1 sequence  
 □ = consensus sequence is supported by 2 sequences  
 ▨ = consensus sequence is supported by 3 sequences  
 ■ = consensus sequence is supported by more than 9 sequences

Fig. 1. Example of a contig assembly. Accession: accession number attributed by GenBank staff after submission of the sequence. Source: first authors of the publication from where the sequence has been retrieved. 5' and 3': direction in which the sequence has been published. Varying bases: mismatching bases where there was no strict majority. At these positions, standard genetic code was used to include every variation. Mismatch: a disagreement between the sequence considered and the majority of the others sequences in the contig

**Confirmation of assembly parameters.** To circumvent the fabrication of chimeric molecules *in silico* and to confirm our assembly parameters, we retrieved from the NCBI website ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)) all ribosomal sequences from 2 widely studied bacteria: *Escherichia coli* and *Salmonella typhimurium*. In total, 136 sequences (72 for *E. coli* and 64 for *S. typhimurium*) were downloaded and assembled into contigs using the SeqManII software. The same assembly parameters as described above were used to cluster this specific dataset to determine if any sequence from *E. coli* was lumped into a contig built from *S. typhimurium* sequences, and vice versa.

**Construction of the MBD and its geographical mapping tool.** The MBD is a MySQL database ([www.mysql.com](http://www.mysql.com)) stored on an Apple X-server ([www.apple.com](http://www.apple.com)). It is fully available on the web at <http://halia.hik.se/mbd>. PHP, a server-side scripting language ([www.php.net](http://www.php.net)), mediates the interactions between the user and the MBD. With a common web browser, any information regarding each sequence, sampling, contig or reference can be extracted from the MBD. A BLAST tool (Altschul et al. 1997) was included on the web server to allow local searches against the current dataset. Submissions of missing metadata or new sequences are possible via a web form. A geographical map tool, programmed in Java ([www.java.com](http://www.java.com)), for global visualization of marine bacterioplankton ribo-

types was linked to the MBD and can be used to display specific query or BLAST results. Note that a recent version of the Java Runtime Engine is necessary to use the mapping tool. Updates of the dataset are performed periodically or upon request for bunch upload of clone libraries or substantial submission to GenBank (e.g. whole genome shotgun sequencing).

**Geographic grouping parameters.** Three main regions (polar, temperate and tropical) were arbitrary delimited to distinguish patterns of distribution (i.e. polar only, polar and temperate, temperate only, temperate and tropical, tropical only and tropical and temperate). Therefore, ribotypes built from only 1 sequence (i.e. singletons), or only 2 sequences, or including less than 3 different sampling sites were discarded for the analysis. In order to evaluate the geographic distribution of each ribotype, contigs were only considered if they included (1) more than 3 sequences; (2) more than 3 sampling sites (1 site = different coordinates and/or different depth, and/or different date [when available], and/or different investigation). Cosmopolitan ribotypes were defined as those including sequences found in at least 3 different sampling sites in the South and/or North poles, North and/or South Temperate regions and Tropical regions. Ribotypes for which no sampling information was provided in GenBank or in the associated reference were not included in the geographic analysis.

**Phylogenetic analysis.** After the contig assembly, consensus sequences were compared and aligned on an approximately 750 bp region using CLUSTAL W for multiple alignment. Sequences that were too short to be aligned and those showing too low similarity to other bacteria were discarded. The phylogenetic tree was then inferred following the Maximum Likelihood method using PAUP (4.0b10) software. Definition names were assigned based on the relevance of the sequences included in each contig: when the sequence belonged to an isolated strain, its name was automatically used; when there were no cultivated members in the contig, the identified members were used; when there were no cultivated or identified members in the contigs, we used the taxonomic definition provided by the authors.

## RESULTS

### SSU rDNAs from marine bacterioplankton submitted to GenBank

On September 8, 2004, we retrieved 7070 SSU rDNA sequences belonging to cultivated and uncultivated free-living marine bacterioplankton cells sampled since 1990. From this initial dataset, we associated sampling coordinates for 3177 SSU rDNA sequences (~44% of the total), which offered a starting point to investigate the distributions of marine bacterioplankton ribotypes. Fig. 1 presents an example of the contig assembly and the related data for each sequence. After assembling the initial set of sequences, we delineated 2352 distinct ribotypes at a  $\geq 97\%$  sequence similarity level, and 2936 distinct ribotypes using  $\geq 98\%$  similarity level. Using the  $\geq 97\%$  similarity level dataset, sampling coordinates were associated to sequences clustered into contigs. These coordinates were then plotted on a world map to visualize the patterns of distribution of the ribotypes. The MBD currently stores 1336 distribution maps of distinct ribotypes based on the  $\geq 97\%$  similarity level dataset. Based on different sampling positions, depths, dates, and/or investigations, we distinguished 41 sampling sites in the polar regions (31 in the northern hemisphere and 10 in the southern hemisphere), 90 sampling sites in the temperate regions (86 in the northern hemisphere and 4 in the southern hemisphere), and 17 sampling sites in the tropical region (Fig. 2).

### Confirmation of assembly parameters

While the contig assembly produces an extensive set of consensus sequences, it can be anticipated that, in some instances, chimeric ribotypes may arise. To evaluate the likelihood of such chimera formation, the

close similarity between *Escherichia coli* and *Salmonella typhimurium* ribosomal genes was considered under identical assembly parameters. From the 136 sequences retrieved for *E. coli* and *S. typhimurium* 16S rDNA, 14 contigs from the 2 ribotypes could be formed with a 97% similarity limit on a >50 bp overlap. This amount of variability was expected since these 2 species are known to carry multiple copies of the 16S rDNA gene. However in 1 contig (data not shown), 3 sequences submitted as *S. typhimurium* were assembled into 1 *E. coli* majority contig. The rest of the contigs were either *E. coli* or *S. typhimurium*, i.e. there was no massive formation of chimeric molecules. If the requirement for matching overlap was set to 98% the 3 'misplaced' *S. typhimurium* sequences shifted to contigs including only *S. typhimurium* sequences, but the total number of contigs increased by 1. These results confirmed that although *E. coli* and *S. typhimurium* are known to carry closely related ribosomal sequences, they were delineated by our approach.

### Patterns of distribution of marine bacterioplankton ribotypes

A summary of the patterns of distribution of the different ribotypes is presented in Fig. 3. A significant portion (10%) of the ribotypes showed ubiquitous distribution, 63% were found in 2 areas (temperate and polar, or temperate and tropical), while 32% were found from only 1 restricted area (polar or temperate). No ribotype in the dataset showed a distribution limited to the tropical region only. The limited number of sampling positions (only 17) from this area most likely justifies this result. Additionally, there was no ribotype with a distribution pattern confined to both polar and tropical regions.

The identity of the contigs presenting specific patterns of distribution is described in Fig. 4. Eleven ribotypes showing ubiquitous distribution belonged to frequently found taxonomic groups of the marine bacterioplankton community (Fig. 5). The  $\alpha$ -proteobacteria were represented by 2 members of the SAR11-clade (*Pelagibacter ubique* contig 346 and uncultured  $\alpha$ -proteobacteria NH46-1 contig 477) and 1 member of the *Roseobacter* group (SCB32 contig 971). The  $\gamma$ -proteobacteria were presented by 1 member of the SAR86-clade (uncultured  $\gamma$ -proteobacteria OM10 contig 269), and 4 isolated strains (*Alteromonas macleodii* contig 11, *Pseudoalteromonas haloplanktis* contig 137, *Marinobacter aquaeoiei* contig 240 and  $\gamma$ -proteobacteria SWAT9 contig 818). Moreover, 1 uncultured  $\delta$ -proteobacteria, the SAR324 (contig 405), 1 strain of the cyanobacteria *Synechococcus* (contig 138) and 1 isolated *Bacteroidetes* bacteria (contig 664) were

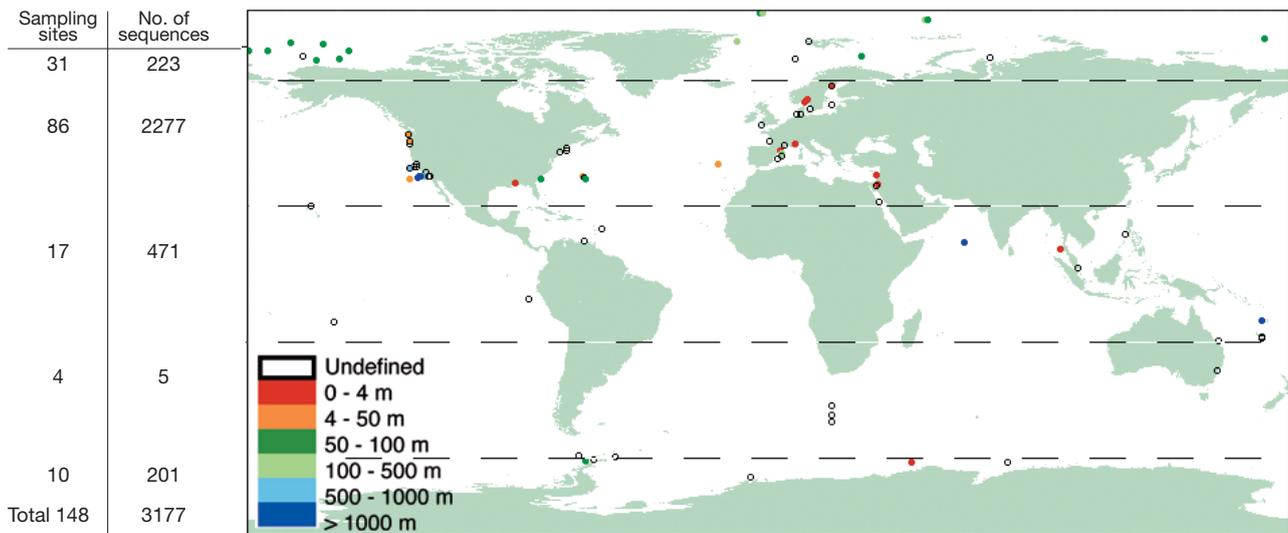


Fig. 2. World map showing all sampling sites retrieved directly from the GenBank records or from their corresponding references

also found with a ubiquitous distribution. Noticeably, we found no evidence for cosmopolitan distribution for any members of the *Firmicutes* group.

Forty-one contigs presented a pattern of distribution confined to temperate and polar regions, which was therefore the most common pattern in our study (Fig. 4). Most ribotypes identified as *Bacteroidetes* followed this pattern of distribution, although one ribotype in this group was cosmopolitan (contig 664), and another showed a distribution limited to temperate and tropical regions (contig 1149). Similarly, ribotypes identified as *Pseudoalteromonas* sp. (contigs 1, 195, 334 and 873) were mainly found in the temperate and polar regions, with the exception of 2 cosmopolitan ribotypes (contigs 137 and 818) and 1 ribotype (contig 2) found in temperate and tropical regions.

Eight ribotypes were found to be confined to the polar regions of the globe, with no occurrence elsewhere, which is remarkable considering the extensive

sampling effort in temperate regions (Fig. 6). All sequences included in these ribotypes were obtained by culture-independent techniques (e.g. cloning or direct amplification of environmental samples). Three of them were  $\gamma$ -proteobacteria (contigs 38, 114, 215), 2 were  $\beta$ -proteobacteria (contigs 560 and 621), 1  $\delta$ -proteobacteria (contig 564), 1 *Bacteroidetes* (contig 665) and 1 an unidentified marine bacterium (contig 1729). A BLAST search (Altschul et al. 1997) for this latter ribotype indicated a *Roseobacter* species as nearest neighbor, but with only 93 % similarity to our query.

### 98 % similarity level for contig assembly

The  $\geq 97$  % similarity limit is a commonly used limit to cluster similar sequences. To investigate the sensitivity of the analysis to this parameter, the dataset was assembled using a  $\geq 98$  % sequence similarity. From the initial 7070 SSU rRNA gene sequences, we could delineate 2936 ribotypes at a  $\geq 98$  % sequence similarity level. The amount of sequences that could be grouped geographically decreased to 1183. Indeed, the number of singleton sequences increased from 769 sequences (at  $\geq 97$  %) to 981 sequences at ( $\geq 98$  %). Moreover, the number of contigs with only 2 sequences or less than 3 sampling sites increased from 298 (at  $\geq 97$  %) to 340 (at  $\geq 98$  %). The patterns of distribution of 9 ribotypes differed from those observed at  $\geq 97$  %. Interestingly, 3 ribotypes (contigs 137, 346 and 664) found to be cosmopolitan in the first assembly showed a different pattern of distribution with a higher similarity level: one was restricted to temperate regions only (contig 664), one was restricted to temperate and polar

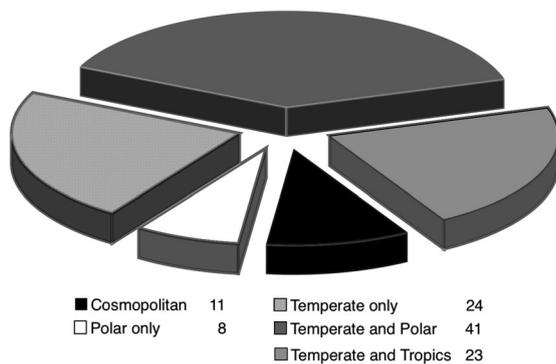


Fig. 3. Overall distribution of all contigs based on their patterns of dispersal

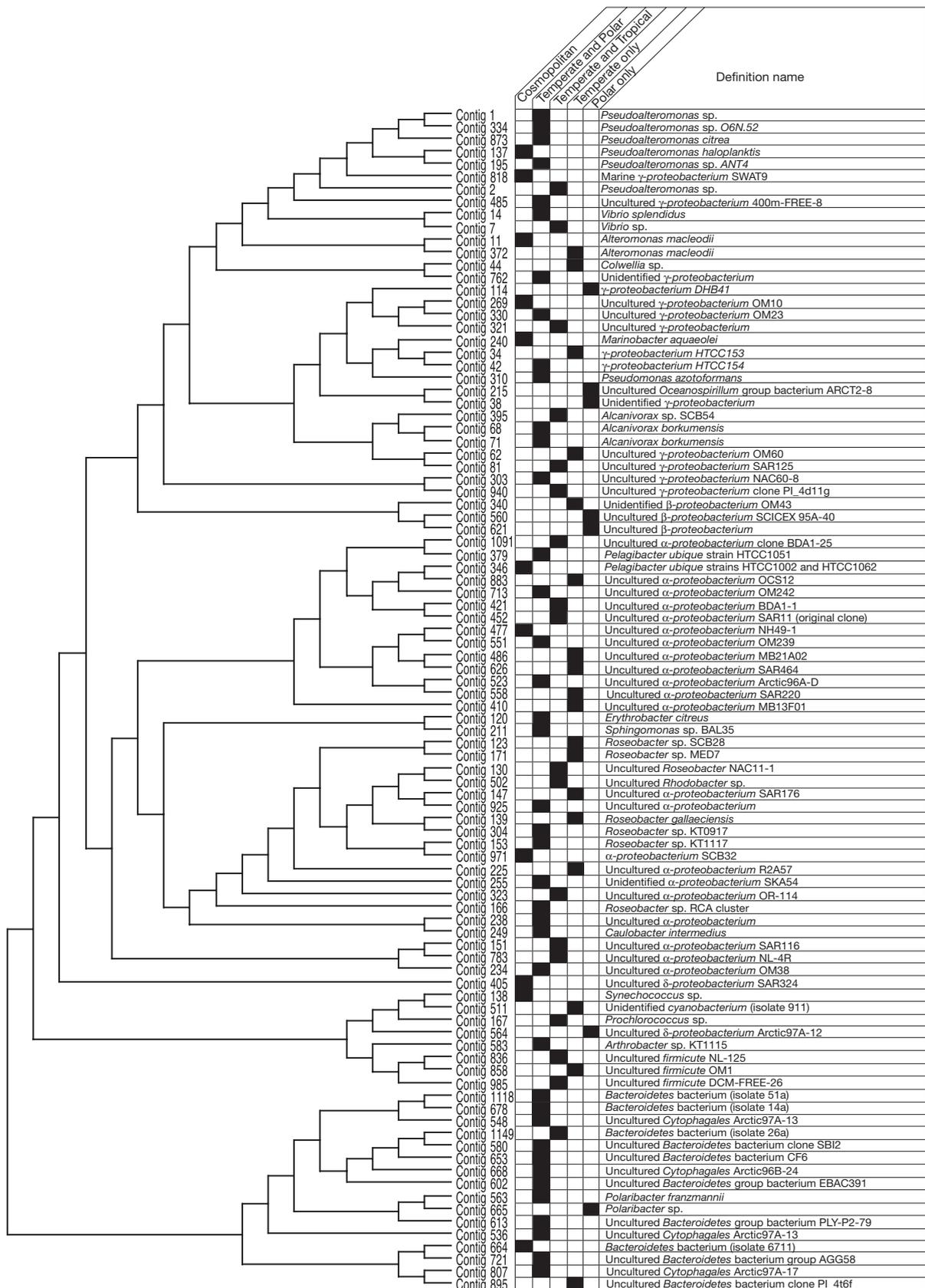
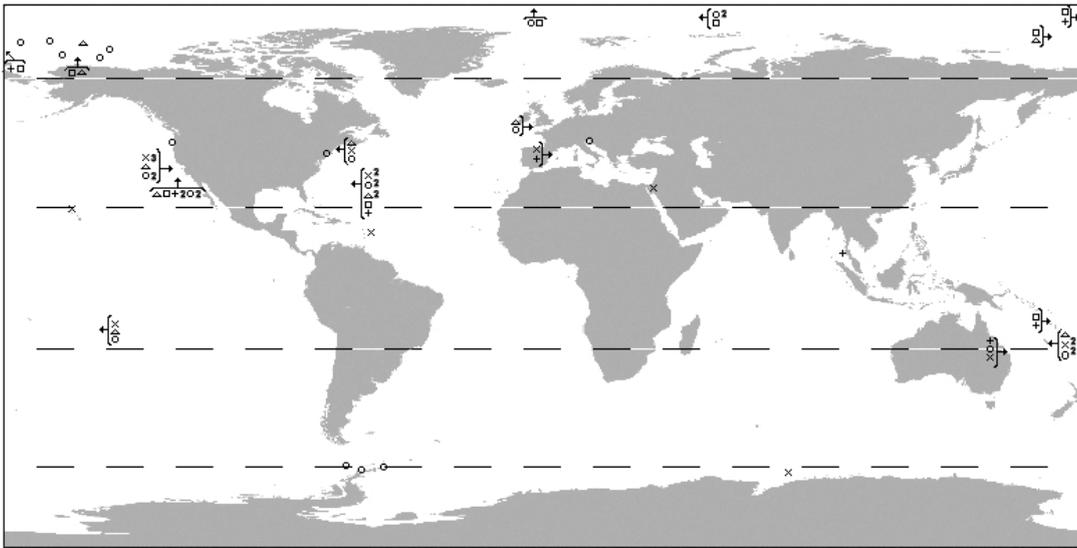


Fig. 4. Phylogenetic tree of contigs with their patterns of distribution. The phylogenetic relationships of the consensus sequences were inferred from an alignment of an approximately 750 bp region of the SSU rRNA gene sequence using CLUSTAL W for multiple alignment and Maximum Likelihood method for construction of the tree using PAUP (4.0b10)



Legend	Contig no.	No. of sequences in the contig	Main representant
x	138	184	<i>Synechococcus</i> sp.
o	346	173	<i>Pelagibacter ubique</i>
	269	87	Uncultured $\gamma$ -proteobacteria OM10
+	11	72	<i>Alteromonas macleodii</i>
Δ	477	49	Uncultured $\alpha$ -proteobacteria NHF49-1
	664	49	Bacteroidetes (isolate 6711)
	137	34	<i>Pseudomonas haloplanktis</i>
□	405	23	Uncultured $\delta$ -proteobacteria SAR324
	240	11	<i>Marinobacter aquaeolei</i>
	971	10	Uncultured $\alpha$ -proteobacteria SCB32
	818	3	Marine $\gamma$ -proteobacteria SWAT9

Fig. 5. World map showing cosmopolitan distribution of some ribotypes

Legend	Contig no.	No. of sequences in the contig	Main representant
x	621	18	Uncultured $\beta$ -proteobacterium
□	38	17	Unidentified $\gamma$ -proteobacterium
Δ	665	12	<i>Polaribacter</i> sp.
◇	215	7	Uncultured <i>Oceanospirillum</i> ARCT2-8
	564	7	Uncultured $\delta$ -proteobacteria ARTIC97
+	560	5	Uncultured $\beta$ -proteobacteria SCICEX95
	114	4	Uncultured $\gamma$ -proteobacteria DHB41
*	1729	4	Unidentified marine bacterium

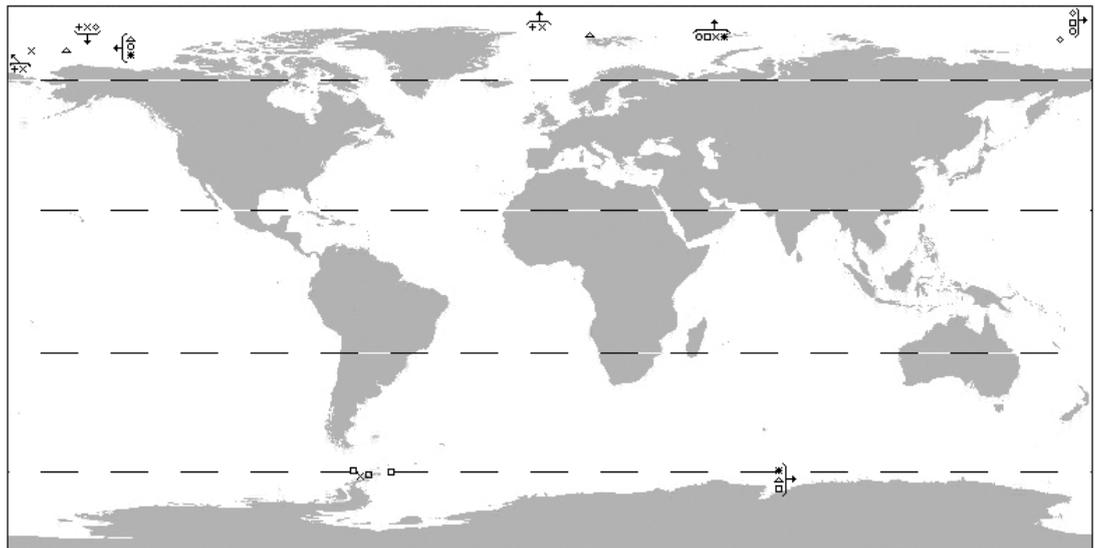


Fig. 6. World map showing distribution restricted to polar regions of some ribotypes

regions (contig 137), and the last one was restricted to temperate and tropical regions (contig 346). The 6 other ribotypes were found to be restricted to temperate regions when assembled at  $\geq 98\%$  similarity, while at a  $\geq 97\%$  sequence similarity they were found either restricted to temperate and polar regions or to temperate and tropical regions. In conclusion, the patterns of distribution we found at a  $\geq 97\%$  sequence similarity level were not drastically changed by increasing the similarity level to  $\geq 98\%$ .

## DISCUSSION

Our analysis showed that the bacterioplankton ribotype records collected on public databases for the past 2 decades can illustrate global distributions of several important members of the global marine bacterioplankton community. The SSU rDNA gene has been extensively used to detect organisms that challenge cultivation techniques (Giovannoni & Rappé 2000). As sequencing has become more common and affordable, recent efforts to build large clone libraries (Acinas et al. 2004, O'Sullivan et al. 2004) have significantly contributed to building a detailed picture of the bacterial components of the marine environment. Lately the interest in the microbial community of the ocean has reached a new level with the use of environmental shotgun sequencing (Venter et al. 2004). One significant result of these studies has been a 3-fold increase in the number of SSU rRNA sequences submitted to GenBank over the past 4 yr. The contribution of the environmental shotgun sequencing from the Sargasso Sea can be observed in our assembly in 153 contigs; of these, 95 contigs included sequences only retrieved by this attempt to sequence the whole bacterioplankton community.

Compared with our first attempt to estimate marine bacterioplankton diversity from SSU rRNA sequences submitted to GenBank (Hagström et al. 2002), the continuous inflow of information to the public databases was illustrated by more than a doubling of the total number of retrieved sequences (7070 vs. 2736 sequences) as well as the number of distinct ribotypes after assembly (2352 vs. 1117 contigs). This increase seemingly contradicts our previous suggestion that 'much of the marine microbial species richness has been sampled' (Hagström et al. 2002); however, the total number of contigs remains surprisingly low given the intense sampling effort. The constant input of unique sequences (~50% of all reported sequences) may, however, point to accumulating neutral mutations that may result in the emergence of new ecotypes after periodic selection (Cohan 2002, Thompson et al. 2005).

Distribution patterns are fundamental in ecology because of their implications for key biological indices, such as species diversity or degree of community resilience. Concurrent to the opening of the 'microbial black box' with the advance of culture-independent identifications, several workers have begun to directly investigate the question of whether free-living bacteria are cosmopolitan or endemic. On the one hand, reports of cosmopolitan distribution were mainly focused on particular bacterioplankton groups (Massana et al. 2000, Morris et al. 2002, Ward & O'Mullan 2002, O'Sullivan et al. 2004, Abell & Bowman 2005), and reflect the occurrence of dominant ribotypes. On the other hand, reports of endemism were often in relation to extremophiles, since they failed to disperse viable forms through shifting environment (Papke et al. 2003, Whitaker et al. 2003) or in relation to bacteria associated to particular organisms (e.g. coral mats [Knowlton & Rohwer 2003] or sponges [Taylor et al. 2005]), which lack a specific host. Our study aimed to provide a global perspective on all members of the free-living marine bacterioplankton community.

The contig assembly approach used here confirmed the ubiquity of some groups (e.g. *Pelagibacter ubique*, *Synechococcus*) as well as earlier reports of exclusion from polar regions (Partensky et al. 1999) of *Prochlorococcus* sp. (contig 167) or the confined pattern of distribution to temperate and polar regions of the *Roseobacter* RCA cluster (Selje et al. 2004) (contig 166). It is therefore tempting to suggest analogous patterns of distribution for some members of the bacterioplankton community.

Previously characterized bacterioplankton, such as *Alteromonas macleodii*, *Pseudoalteromonas haloplanktis*, *Marinobacter aquaeolei* and a member of the SAR86 cluster, surely present ubiquitous distribution. In particular, *A. macleodii* has been shown to be highly competitive when associated to aggregate matter (Long & Azam 2001), and this trait may explain the ubiquitous occurrence of this microorganism. In contrast to cosmopolitan distribution, the current dataset also supported the confinement of specific ribotypes to the polar regions of the globe. Despite continuous sampling in the temperate and tropical areas, 8 taxa were never found outside the polar regions and may be regarded as endemic to cold environments. Selective adaptation to constantly cold conditions may preclude successful establishment in temperate areas where high temperatures purge the environment annually.

Even though our data could not illustrate the pattern of distribution limited to tropical areas, future intensive sampling in these regions might also show similar traits for the tropical bacterioplankton community. Similarly, some ribotypes may be seen as restricted to temperate regions of the globe, but since >70% of the

sequences presenting sampling coordinates were found in the temperate regions, it seems likely that our dataset was biased for this pattern. Therefore, it may be speculative to describe the occurrence of certain ribotypes restricted to temperate regions, although this cannot be excluded. With additional sampling data, especially from the tropics and the polar regions, some ribotypes may show different patterns of distribution, ranging from true endemism (i.e. occurring at only 1 location) to cosmopolitanism.

The particular case of the *Roseobacter* subgroup of the  $\alpha$ -*proteobacteria* is interesting. As previously shown by Selje et al. (2004), our approach confirmed that 1 ribotype (contig 166) was confined to temperate and polar regions of the globe. With the current data, 3 other *Roseobacter* ribotypes (contigs 153 and 238, 249) showed similar patterns of distribution. However, 5 *Roseobacter* ribotypes (contigs 123, 139, 147, 171 and 225) were found only in temperate regions, whereas 6 were limited to temperate and tropical regions (contigs 130, 255, 304, 323, 502 and 925). As seen previously for *Prochlorococcus* strains (West et al. 2001), different members of the *Roseobacter* clade may show different patterns of distribution.

Cosmopolitan distribution of some members of the SAR11 clade has been previously shown (Morris et al. 2002). Despite the ubiquitous distributions of 2 ribotypes belonging to the SAR11 cluster (contigs 346 and 446), our approach distinguished 3 ribotypes that were also found with a distribution confined to temperate and tropical regions (contigs 421, 452 and 1091), and 4 ribotypes confined to temperate and polar regions (contigs 379, 523, 551 and 713). This result suggests several occurrences of geographic niche separation among ribotypes belonging to the SAR11 clade, which may be analogous to the depth separation suggested for some members of this clade (Field et al. 1997).

Similarly, several ribotypes reported as *Pseudoalteromonas* spp. presented different patterns of distribution. In conclusion, the marine bacterioplankton community certainly includes different taxa showing different patterns of distribution, from strictly endemic to one environment to true cosmopolitanism.

The  $\geq 97$  or  $\geq 98$  % similarity limits used here are in concordance with a common rule-of-thumb to delineate organisms that are  $>3$  % different on the SSU rDNA sequence into different species. Basically, this rule implies that speciation is proportional to evolutionary distance when used to cluster 16S rRNA gene sequences into consensus ribotypes. In a strict phylogenetic sense, the contig assembly lacks theoretical support, but as demonstrated here, it undoubtedly provides valuable information into biogeography of similar ribotypes. Many authors have argued the necessity of adjusting the  $\geq 97$  % similarity limit, or including the

internally transcribed spacer region of the ribosomal operon to delineate very closely related taxa (García-Martínez & Rodríguez-Valera 2000, Suzuki et al. 2001, Roca et al. 2003). Yet, the SSU rDNA genes are used here in a biogeography perspective as trace markers of similar ribosomal genes. Interestingly, most of the patterns of distribution found at  $\geq 97$  % were not affected by restricting the similarity level to  $\geq 98$  %. However, a sharper picture of distribution was observed for 3 ribotypes showing ubiquitous distribution at lower similarity level. Similarly, stringent species delineation methods (e.g. DNA sequence comparisons) could show that some accepted approaches to delineating species of protists (e.g. morphogenesis) combined different taxonomic units (Gonzalez et al. 2001). It is therefore not surprising to observe more accurate patterns of distribution with more constrained parameters for phylogenetic delineation.

The contig assembly approach was used here to group similar SSU rDNA gene sequences obtained from different sampling sites/times and display the resulting patterns of distribution. Assembling partial sequences into clusters to reduce redundancy is analogous to whole-genome shotgun (WGS) sequencing, which is the common strategy in genomic projects (Venter et al. 2003). Moreover, this method corresponds to the strategy of the 'bacterial artificial chromosome project' that also depicts the marine bacterial genetic diversity, i.e. to sequence a massive amount of genomic DNA to reveal the identity of anonymous bacterioplankton species (Beja et al. 2000, 2002). Compared to conventional phylogenetic analysis (i.e. gene cluster tree), this method benefits from the additional information provided by longer ribosomal sequences from similar ribotypes. Hence, shorter fragments scattered along the SSU rDNA gene can be aligned and grouped in clusters of similar genes. Potential sequencing errors may thus be corrected by the recurrent addition of new sequences in the assembly. Nevertheless, relevant fine-scale phylogenetic differences may also be masked in the resulting consensus sequences.

As shown in this study, the large amount of sequence information stored on public databases (e.g. GenBank) is valuable to the performing of biogeography studies. However, because too few data were directly available from GenBank records or their associated literature, sampling depth or water temperature (obvious factors influencing aquatic biogeography) could not be considered. However, with the addition of sampling site information and of relevant environmental metadata (e.g. water temperature, bacterial count, primary production), ecological studies with a truly large-scale perspective focused on, for example, community structure and biodiversity would certainly be possible.

Interested authors are encouraged to visit our website (<http://halia.hik.se/mbd>) and add geo-references to their sequences.

The widespread occurrence of some members of the marine bacterioplankton community raises questions regarding the processes of their dispersion. The current literature contains few studies on the natural dispersal of components of the bacterioplankton. Whereas major ocean currents and water mass movements would seem to favor global mixing of microorganisms, these transport processes are slow relative to seasonal variation and growth rates of bacteria. However, bacterioplankton is constantly transferred from water to air through the bursting of air bubbles produced in breaking waves (Blanchard & Syzdek 1970). The atmospheric flow thus constitutes a mechanism for the dispersal of viral particles as well as bacteria (Baylor et al. 1977, Griffin et al. 2001). Disclosing processes that generate the different patterns of distribution found here thus appear to be a worthwhile challenge for microbial biogeography.

**Acknowledgements.** This work was supported by the Natural Science Faculty of Kalmar University and the Swedish Research Council, Grant No. 621-2001-2814. The Apple X-server hosting the MBD was financed through the EU Project 'BASICS' contract EVK3-CT-2002-00078. We thank the Devment Group (Kalmar, Sweden) for their support in developing the graphical map interface.

#### LITERATURE CITED

- Abell GCJ, Bowman JP (2005) Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiol Ecol* 51:265–277
- Acinas SG, Anton J, Rodríguez-Valera F (1999) Diversity of free-living and attached bacteria in offshore Western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Appl Environ Microbiol* 65:514–522
- Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 430:551–554
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Azam F (1998) Microbial control of oceanic carbon flux: the plot thickens. *Science* 280:694–696
- Baylor ER, Baylor MB, Blanchard DC, Syzdek LD, Appel C (1977) Virus transfer from surf to wind. *Science* 198: 575–580
- Beja O, Suzuki MT, Koonin EV, Aravind L and 8 others (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* 2:516–529
- Beja O, Koonin EV, Aravind L, Taylor LT and 6 others (2002) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl Environ Microbiol* 68:335–345
- Blanchard DC, Syzdek L (1970) Mechanism for the water-to-air transfer and concentration of bacteria. *Science* 170:626–628
- Brock TD (1961) *Milestones in microbiology*. Prentice-Hall, Englewood Cliffs, NJ
- Carr NG, Mann NH (1994) The oceanic cyanobacterial picoplankton. In: Bryant DA (ed) *The molecular biology of Cyanobacteria*. Kluwer Academic Publishers, Dordrecht, p 27–48
- Cho JC, Tiedje JM (2000) Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl Environ Microbiol* 66:5448–5456
- Cohan FM (2002) What are bacterial species? *Annu Rev Microbiol* 56:457–487
- Coleman AW (2001) Biogeography and speciation in the *Pandorina volvulina* (Chlorophyta) superclade. *J Phycol* 37: 836–851
- Cottrell MT, Kirchman DL (2000) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Appl Environ Microbiol* 66:5116–5122
- Darling KF, Wade CM, Stewart IA, Kroon D, Dingle R, Brown AJ (2000) Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature* 405:43–47
- DeLong EF (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci USA* 89:5685–5689
- Ducklow HW (2000) Bacterial production and biomass in the oceans. In: Kirchman DL (ed) *Microbial ecology of the oceans*. Wiley-Liss, New York, p 85–120
- Fenchel T, Finlay BJ (2004) The ubiquity of small species: patterns of local and global diversity. *Bioscience* 54:777–784
- Field KG, Gordon D, Wright T, Rappe M, Urback E, Vergin K, Giovannoni SJ (1997) Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* 63:63–70
- Finlay BJ (1998) The global diversity of protozoa and other small species. *Int J Parasitol* 28:29–48
- Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. *Science* 296:1061–1063
- Finlay BJ, Clarke KJ (1999) Apparent global ubiquity of species in the protist genus *Paraphysomonas*. *Protist* 150: 419–430
- Fuhrman JA, McCallum K, Davis AA (1993) Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. *Appl Environ Microbiol* 59:1294–1302
- García-Martínez J, Rodríguez-Valera F (2000) Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I. *Mol Ecol* 9:935–948
- Giovannoni SJ, Rappé MS (2000) Evolution, diversity and molecular ecology of marine prokaryotes. In: Kirchman DL (ed) *Microbial ecology of the oceans*. Wiley-Liss, New York, p 47–84
- Gonzalez MA, Coleman AW, Gomez PI, Montoya R (2001) Phylogenetic relationship among various strains of *Dunaliella* (chlorophyceae) based on nuclear ITS rDNA sequences. *J Phycol* 37:604–611
- Griffin DW, Garrison VH, Herman JR, Shinn EA (2001) African desert dust in the Caribbean atmosphere: microbiology and public health. *Aerobiologia* 17:203–213
- Hagström Å, Pinhassi J, Zweifel UL (2000) Biogeographical diversity among marine bacterioplankton. *Aquat Microb Ecol* 21:231–244
- Hagström Å, Pommier T, Rohwer F, Simu K, Stolte W, Svensson D, Zweifel UL (2002) Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl Environ Microbiol* 68:3628–3633
- Hollibaugh JT, Bano N, Ducklow HW (2002) Widespread distribution in Polar oceans of a 16S rRNA gene sequence

- with affinity to *Nitrospira*-like ammonia-oxidizing bacteria. *Appl Environ Microbiol* 68:1478–1484
- Knowlton N, Rohwer F (2003) Multispecies microbial mutualisms on coral reefs: the host as a habitat. *Am Nat* 162:S51–S62
- Long RA, Azam F (2001) Antagonistic interactions among marine pelagic bacteria. *Appl Environ Microbiol* 67:4975–4983
- Massana R, DeLong EF, Pedros-Alio C (2000) A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol* 66:1777–1787
- Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420:806–810
- Mullins TD, Britschgi TB, Krest RL, Giovannoni SJ (1995) Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. *Limnol Oceanogr* 40:148–158
- O'Sullivan LA, Fuller KE, Thomas EM, Turley CM, Fry JC, Weightman AJ (2004) Distribution and culturability of the uncultivated 'AGG58 cluster' of the Bacteroidetes phylum in aquatic environments. *FEMS Microbiol Ecol* 47:359–370
- Papke RT, Ramsing NB, Bateson MM, Ward DM (2003) Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* 5:650–659
- Partensky F, Hess WR, Vault D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63:106–127
- Rocap G, Larimer FW, Lamerdin J, Malfatti S and 20 others (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042–1047
- Selje N, Simon M, Brinkhoff T (2004) A newly discovered Roseobacter cluster in temperate and polar oceans. *Nature* 427:445–448
- Staden R (1979) A strategy of DNA sequencing employing computer programs. *Nucleic Acids Res* 6:2601–2610
- Staley JT, Gosink JJ (1999) Poles apart: biodiversity and biogeography of sea ice bacteria. *Annu Rev Microbiol* 53:189–215
- Suzuki MT, Beja O, Taylor LT, DeLong EF (2001) Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. *Environ Microbiol* 3:323–331
- Taylor MW, Schupp PJ, de Nys R, Kjelleberg S, Steinberg PD (2005) Biogeography of bacteria associated with the marine sponge *Cymbastela concentrica*. *Environ Microbiol* 7:419–433
- Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V and 5 others (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* 307:1311–1313
- Venter JC, Levy S, Stockwell T, Remington K, Halpern A (2003) Massive parallelism, randomness and genomic advances. *Nat Genet* 33(Suppl):219–227
- Venter JC, Remington K, Heidelberg JF, Halpern AL and 19 others (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74
- Ward BB, O'Mullan GD (2002) Worldwide distribution of *Nitrosococcus oceani*, a marine ammonia-oxidizing gamma-proteobacterium, detected by PCR and sequencing of 16S rRNA and amoA genes. *Appl Environ Microbiol* 68:4153–4157
- Waterbury JB, Watson SW, Guillard RRL, Brand LE (1979) Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277:293–294
- West NJ, Schonhuber WA, Fuller NJ, Amann RI, Rippka R, Post AF, Scanlan DJ (2001) Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* 147:1731–1744
- Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301:976–978

Editorial responsibility: John Dolan,  
Villefranche-sur-Mer, France

Submitted: August 29, 2005; Accepted: September 23, 2005  
Proofs received from author(s): October 26, 2005