

16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the Green Non-Sulfur bacteria

(molecular ecology/phylogeny/thermophily)

STEPHEN J. GIOVANNONI*, MICHAEL S. RAPPÉ, KEVIN L. VERGIN, AND NANCI L. ADAIR

Department of Microbiology, Oregon State University, Corvallis, OR 97331

Communicated by Lynn Margulis, University of Massachusetts, Amherst, MA, April 1, 1996 (received for review August 29, 1994)

ABSTRACT Microorganisms play an important role in the biogeochemistry of the ocean surface layer, but spatial and temporal structures in the distributions of specific bacterioplankton species are largely unexplored, with the exceptions of those organisms that can be detected by either autofluorescence or culture methods. The use of rRNA genes as genetic markers provides a tool by which patterns in the growth, distribution, and activity of abundant bacterioplankton species can be studied regardless of the ease with which they can be cultured. Here we report an unusual cluster of related 16S rRNA genes (SAR202, SAR263, SAR279, SAR287, SAR293, SAR307) cloned from seawater collected at 250 m in the Sargasso Sea in August 1991, when the water column was highly stratified and the deep chlorophyll maximum was located at a depth of 120 m. Phylogenetic analysis and an unusual 15-bp deletion confirmed that the genes were related to the Green Non-Sulfur phylum of the domain Bacteria. This is the first evidence that representatives of this phylum occur in the open ocean. Oligonucleotide probes were used to examine the distribution of the SAR202 gene cluster in vertical profiles (0–250 m) from the Atlantic and Pacific Oceans, and in discrete (monthly) time series (0 and 200 m) over 30 consecutive months in the Western Sargasso Sea. The data provide robust statistical support for the conclusion that the SAR202 gene cluster is proportionately most abundant at the lower boundary of the deep chlorophyll maximum ($P = 2.33 \times 10^{-5}$). These results suggest that previously unsuspected stratification of microbial populations may be a significant factor in the ecology of the ocean surface layer.

The ocean surface layer, defined as the region between the surface and the deepest annual extent of wind-driven mixing (ca. 0–250 m), plays a remarkably important role in oceanic biogeochemistry. In temperate oceans, this region is typically characterized by significant gradients in density, light field, temperature, and the concentration of reduced compounds of nitrogen. Furthermore, the photic zone spans the upper region of the surface layer, whereas in the lower surface layer, the flux of sinking particulate organic carbon declines rapidly due to organic carbon remineralization by microbial respiration (1, 2). A recent study of the North Pacific subtropical gyre linked variations in the efficiency of carbon remineralization in the surface layer to long-term changes in upper-ocean mixing brought about by the 1991–1992 El Niño Southern Oscillation event, thereby suggesting that physical stratification of the surface layer might be a factor in the global carbon cycle (3). Although it is recognized that bacterioplankton play a key role in the carbon cycle, little is known about the vertical organization of bacterioplankton communities in the surface layer, with the exception of a few cyanobacterial species (2, 4, 5).

Molecular biology provides new information on bacterioplankton microbial ecology by targeting those species that are

difficult to cultivate. Generally, the strategy has been to clone rRNA genes from environmental samples and, thereby, to substitute gene diversity for microbial diversity (6–9). Although the biases inherent in these methods are not completely understood, the results have been striking. The various studies embracing this strategy have reached a similar conclusion—that previously undescribed organisms provide the most abundant 16S rRNA lineages recovered from natural ecosystems by gene cloning. In addition to revealing new organisms, rRNA sequencing studies provide data for the design of oligonucleotide probes, which can be used to examine the distributions and growth state of specific species in environmental samples (10–13).

16S rDNAs previously isolated from natural bacterioplankton samples are diverse but predominantly fall into several distinct phylogenetic groups, many of which have been recovered repeatedly from different oceans (14, 15). A majority of the gene clones recovered belong to the cyanobacteria and proteobacteria phyla, although Gram-positive bacteria, flavobacteria, species related to *Fibrobacter*, and archaeal genes have also been found. Most samples studied have come from the surface of subtropical regions of the ocean, although samples from 100 m, 500 m, and marine snow have also been examined (16–18).

We obtained the gene clones described in this study from samples collected at the Bermuda Atlantic Time-Series Study site (BATS), which is located in an unusually oligotrophic region of the subtropical Atlantic Ocean (19). The water depth at BATS is about 4000 m. Throughout most of the year, the surface layer at BATS is physically stratified due to variations in salinity and temperature. Stratification increases in the spring and summer due to seasonal changes in day length and a decrease in wind-driven mixing.

We first observed the unusual bacterial genes described in this report by random sequencing of 16S rRNA clones obtained from 250 m in the Sargasso Sea. The sample was part of a series collected in August 1991 from depths between the surface and 250 m. At that time, the water column was highly stratified and the deep chlorophyll maximum (DCM) was located at 120 m. The construction of this clone library was motivated by probe hybridization experiments which suggested that many gene lineages cloned previously from surface samples were rare in the lower region of the surface layer. This suggested that microbial communities in the water column at BATS are stratified and that a unique community of organisms might occur in the lower surface layer.

The gene sequences we describe in this paper are related to the Green Non-Sulfur (GNS) phylum, which is one of the 11

Abbreviations: BATS, Bermuda Atlantic Time-Series Study site; DCM, deep chlorophyll maximum; HMW, high molecular weight; GNS, Green Non-Sulfur.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base [accession nos. U20797 (SAR202), U21682 (SAR263), and U20798 (SAR307)].

*To whom reprint requests should be addressed. e-mail: giovanns@bcc.orst.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

bacterial phyla originally described by Woese and colleagues based on 16S rRNA sequence comparisons (20). This phylum includes only four genera that have been cultivated, *Chloroflexus*, *Heliothrix*, *Herpetosiphon*, and *Thermomicrobium*. All four of these genera contain at least some thermophilic species. *Chloroflexus* and *Heliothrix* are phototrophic or chemoorganotrophic. *Herpetosiphon* and *Thermomicrobium* are respiratory chemoorganotrophs. This phylum occupies a particularly significant position in the phylogeny of bacteria because it appears to have diverged before the main radiation of bacteria, which includes 9 of the 11 phyla first described by Woese (21). Only three genera, *Thermotoga*, *Geotoga*, and *Aquifex*, consistently diverge earlier in phylogenetic trees (10).

METHODS

Sampling. Niskin bottles attached to a CTD (conductivity, temperature, depth) rosette were used to collect water samples from the BATS (31° 50' N, 64° 10' W) and a site in the Pacific Ocean located *ca.* 70 km from the Oregon coast (46° 45' N, 126° 2' W). A CTD is a device that automatically records chemical and physical data as it is lowered by cable into water. In this case, a Sea-Bird CTD was used to measure continuous profiles of temperature and fluorescence. Cells were collected from seawater samples on 0.2- μ m polysulfone filters (Supor-200; Gelman). Water (12–24 liters) was filtered from each depth on four 47-mm filters at 7 psi vacuum. The nominal depths sampled were 0, 40, 80, 120, 160, 200, and 250 m. The filters were then placed in heat-sealing plastic bags with 5 ml of lysis buffer (20 mM EDTA/400 mM NaCl/0.75 M sucrose/50 mM Tris-HCl, pH 9.0) and stored at –20°C or in liquid N₂ for later processing on shore.

Nucleic Acid Extraction. Total cellular nucleic acids were extracted from the filters using a combination of procedures optimized for small sample sizes. To lyse cells, the samples were thawed on ice, SDS was added to 1%, and proteinase K was added to 100 μ g/ml. The bags were resealed, and samples were incubated in an HB-1 hybridization oven (Techne, Cambridge, U.K.) at 37°C for 30 min, then 55°C for 10 min. Lysate was transferred to polypropylene centrifuge tubes, extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1), and extracted again with an equal volume of chloroform/isoamyl alcohol (24:1). The purified nucleic acids were then precipitated with 2 volumes 100% ethanol and 0.1 volume 2.0 M sodium acetate, pH 5.2, and resuspended in aqueous solution. High molecular weight (HMW) RNA and DNA were fractionated by isopycnic centrifugation in cesium trifluoroacetate (1.6 g/cm³). Samples were centrifuged in a 1.4-ml volume for 48 hr, at 40,000 rpm, 15°C in a TLS-55 rotor (Beckman). After centrifugation, DNA was precipitated from the upper 1.3 ml, and RNA was precipitated from the lower 0.1 ml, by the addition of equal volumes of isopropanol. Nucleic acids were then resuspended in 10 mM Tris/1 mM EDTA (TE) buffer and precipitated with 0.1 volume 2.0 M sodium acetate, pH 5.2, and 2 volumes 100% ethanol. DNA was resuspended in TE buffer, and RNA was resuspended in TE/0.5% glutaraldehyde.

Gene Cloning, Sequencing, and Phylogenetic Analyses. Prokaryotic 16S rRNA genes were amplified from environmental DNA for cloning by PCR with *Taq* polymerase (Promega) and bacterial 16S rDNA primers (27F, AGAGTTTGATCMTG-GCTCAG; 1518R, AAGGAGGTGATCCANCCRCA). The reaction volume of 100 μ l contained 10 ng of template, 5% acetamide, each primer at 200 nM, 1.5 mM MgCl₂, 200 μ M nucleotide triphosphates (total), *Taq* buffer, and 2.5 units *Taq* DNA polymerase. Amplifications were performed in a thermocycler (MJ Research, Cambridge, MA) using the following conditions: 30 cycles, annealing for 1 min at 50°C, elongation at 72°C for 3 min, and denaturation at 96°C for 1 min. A single product band of the predicted length was observed after

agarose gel electrophoresis. The clone library was constructed by ligation of the amplification products into the plasmid vector pCRII (Invitrogen), following the manufacturer's instructions. Double-stranded plasmid DNA sequencing was performed using dye-terminator chemistry and an Applied Biosystems 373A automated sequencer, except in the case of clone SAR202. This clone was sequenced both by conventional Sanger dideoxy-terminated sequencing with Sequenase (United States Biochemical) and α -³⁵S-labeled dATP and the automated sequencing method (22). All sequences were determined bidirectionally. Sequence analysis was performed using the programs GDE, supplied by Steve Smith (Millipore, Bedford, MA), and GRNAID, supplied by Shannon Whitmore (Mentor Graphics, Wilsonville, OR). Phylogenetic analyses were restricted to regions of unambiguous nucleotide alignment. All sequences were submitted to the Ribosomal RNA Database Project program CHECK_CHIMERA to detect chimeric artifacts (23). Phylogenetic relationships were inferred by the neighbor-joining method using the Phylogeny Inference Package (PHYLIP) version 3.4 (24). Phylogenetic trees were edited using the program TREETOOL provided by Mike Maciukenas (University of Illinois, Urbana, IL).

Accession Numbers. For the accession numbers of the nucleotide sequences filed in GenBank, see the data deposition footnote. In cases where the sequences of two clones were identical (see Fig. 3), only one was submitted to GenBank.

DNA Vertical Profiles. The 1492R (GGTTACCTTGT-TACGACTT) and 27F primers were used to amplify bacterial rDNAs for dot blots. Reaction conditions were as given above with the following changes: 35 cycles, annealing for 1 min at 55°C, elongation at 72°C for 3 min, and denaturation at 94°C for 1 min. Amplicons were purified with a Qiaquick-spin PCR purification kit (Qiagen, Chatsworth, CA).

PCR products were resuspended in 180 μ l of TE buffer. Products were denatured by the addition of 20 μ l of 2.0 M NaOH and incubated 10 min at room temperature before blotting on Zetaprobe nylon membrane (Bio-Rad). Each well was washed with 200 μ l of 2 \times standard saline phosphate/EDTA (SSPE; 1 \times SSPE: 0.18 M NaCl/10 mM NaPO₄, pH 7.2/1 mM EDTA) to neutralize NaOH. The amounts of amplified DNA blotted were 100 ng per well from natural populations and 30 ng per well for controls. DNA concentrations were determined by optical absorbance at 260 nm. Membranes were dried under vacuum at 70–80°C for 15 min, exposed to 120 mJ/m² of 260 nm UV radiation, and stored desiccated before probing. Results were expressed as the percent of amplicons according to the following equation:

$$\Phi_x = \left[\left(\frac{P_x}{P_c} \right) \cdot \left(\frac{P'_x}{P'_c} \right)^{-1} \right] \times 100$$

where Φ_x is the percentage of SAR202B genes in the mixture of amplicons, P_x is the hybridization of the SAR202B probe to the environmental amplicons (representing all bacterial 16S rRNA genes in the community), P_c is the hybridization of the 338R probe to environmental amplicons, and the primed variables are values for hybridizations to a homogenous preparation of amplicons of the SAR202 gene.

RNA Vertical Profiles. RNA concentrations were determined by optical absorbance at 260 nm. Portions (100, 50, 20, and 10 ng) of each RNA sample were resuspended in TE/0.5% glutaraldehyde, heated to 37°C for 2 min, and blotted using a dot blot manifold (Minifold I SRC 096/90; Schleicher & Schuell) onto Zetaprobe nylon membrane. RNA was crosslinked to filters by UV irradiation as described above. The hybridization values for HMW RNA (see Figs. 4 and 5) are expressed in relative units, as the ratio of specific probe hybridization to bacterial 338R probe hybridization.

Oligonucleotide Probes. Two oligonucleotide probes were designed to complement the 16S rRNA of SAR202 and related species. Probe SAR202AR (AAGACGCGAGCCGCTCCTCCG) was originally designed from the SAR202 gene sequence alone; it corresponds to positions 222–242 of the *Escherichia coli* 16S rRNA, was used at low hybridization stringency to identify further clones of the same type in the library, and mismatches the SAR307 gene, which was discovered later by random sequencing, at two positions. SAR202BR (AGGGTGTGTTACTCACGC) corresponds to *E. coli* positions 122–138 and matches the sequences of both SAR202 and SAR307. These sites were chosen as targets for probes because they are only moderately variable in prokaryotes. A probe (SAR6R) was also designed to hybridize to *Prochlorococcus* (TCCTCAAGTACCGTCATA). It corresponds to *E. coli* positions 467–495. Hybridization stringencies for all probes were determined empirically by washing replicate blots at different temperatures.

T4 polynucleotide kinase was used to label the 5' terminus of each of the oligonucleotide probes (25). An oligonucleotide that is specific for the domain Bacteria (338R, GCTGCCTCCCGTAGGAGT) was used to evaluate the amounts of bacterial rDNA on blots.

Blots were prehybridized in 15 ml Z-hyb buffer (1 mM EDTA/0.5 M NaH₂PO₄, pH 7.2/7% SDS) for 10–30 min at room temperature, the prehybridization buffer was decanted, and 6 ml of fresh Z-hyb buffer containing 20–50 ng of oligonucleotide was added. Blots were hybridized at 30°C for 5 hr to overnight and were washed three times for 15 min at room temperature and one time for 15 min at the stringency temperature in 25 ml of 0.2× SSPE/0.1% SDS. Stringent wash conditions were 37°C for the SAR202AR probe and 45°C for both the 338R probe and the SAR6R probe. The SAR202BR probe was washed with 2.0× SSPE/0.1% SDS at a stringency temperature of 40°C.

Washed blots were exposed to PhosphorImager plates and detected using a Molecular Dynamics PhosphorImager SI and IMAGEQUANT software.

RESULTS AND CONCLUSIONS

The unique phylogenetic position of the SAR202 gene cluster was first noticed as a result of random sequencing of a library of 142 clones. The clone library was constructed specifically to search for evidence of novel microorganisms that might occur below the euphotic zone. Eventually, two clones belonging to this gene cluster were sequenced completely (SAR202 and SAR307), and three more (SAR279, SAR293, and SAR263) were sequenced in the domain corresponding to *E. coli* positions 936–1346. All of the genes were related, and collectively constituted a new, deep branching lineage of the GNS phylum.

A phylogenetic analysis of clones SAR202 and SAR307 is presented in Fig. 1. The topology shown was determined by the neighbor-joining method. Analyses using Wagner parsimony, as well as the maximum likelihood method, yielded the same phylogenetic position for SAR202 and SAR307. In every case, the clones formed a monophyletic lineage with other members of the GNS phylum. This relationship was supported by 88% of bootstrap replicates with Wagner parsimony and 97% of bootstrap replicates of neighbor-joining trees (Fig. 1). The inclusion of SAR202 and SAR307 did not affect the position of the GNS phylum with respect to other bacterial phyla, reinforcing the original phylogenetic position determined for this phylum (20).

A secondary structural model for the SAR202 gene product was used to examine the sequence for unusual base pairing in conserved helices (Fig. 2). The model confirmed that the predicted gene product could be folded into a consensus secondary structure that matched the major conserved features of bacterial 16S rRNAs.

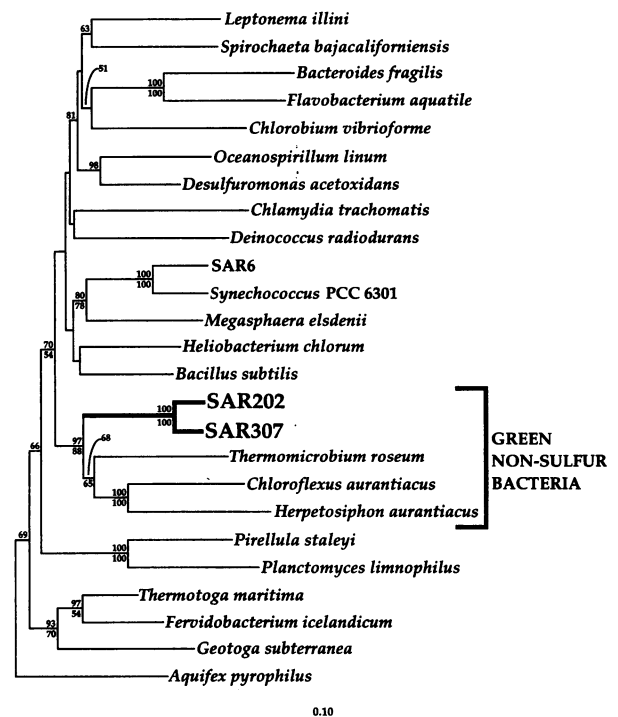


FIG. 1. Phylogenetic tree showing relationships of SAR202 and SAR307 to representative bacterial 16S rRNA genes. The tree was inferred from nearly complete sequences by the neighbor-joining method. The numbers of bootstrap replicates that supported the branching order, from a total of 100 replicates, are shown above (neighbor joining) and below (Wagner parsimony) the internal segments. Values below 50% are not shown. Gene sequences from *Haloferax volcani* and *Methanococcus voltae* were used to determine the root of the tree. The scale bar indicates substitutions per nucleotide position.

The phylogenetic position of the SAR202 gene lineage was confirmed by an analysis of signature sequences. Signature sequences are characteristic nucleotides found at specific positions only in a phylogenetically related group of genes. The numbers cited below refer to the positions of nucleotides in the *E. coli* 16S rRNA, which we use here as a frame of reference. Two very unusual secondary structure signatures of the GNS phylum, a deletion of 15 bases between positions 1123 and 1147 and a loop structure between positions 607 and 630 (Fig. 2), are both present in the SAR202 lineage (20). Woese (21) also described five diagnostic base substitutions that defined the GNS phylum. Three of these, a G residue at position 53, an A residue at position 906, and a G residue at position 1410, are found in the SAR202 gene. Two GNS signatures described by Woese, G residues at positions 1202 and 1224, were instead U residues in SAR202, which are the consensus nucleotides for the Bacteria domain at these positions. This deviation from the signatures described by Woese is consistent with the phylogenetic position of SAR202, which branches much earlier than other members of the phylum. Hence, the G residues at positions 1202 and 1224 should be regarded as derived traits rather than those inherited from the common ancestor of the phylum.

Initially, a single oligonucleotide probe, SAR202AR, was designed to selectively hybridize to clone SAR202. It was used at low hybridization stringency to screen the 250-m clone library for further clones related to SAR202. Ten clones hybridized to the SAR202AR probe. Four of these clones were partially sequenced and identified as members of the same gene cluster. SAR307, also a member of the SAR202 cluster by phylogenetic criteria, was encountered by random sequenc-

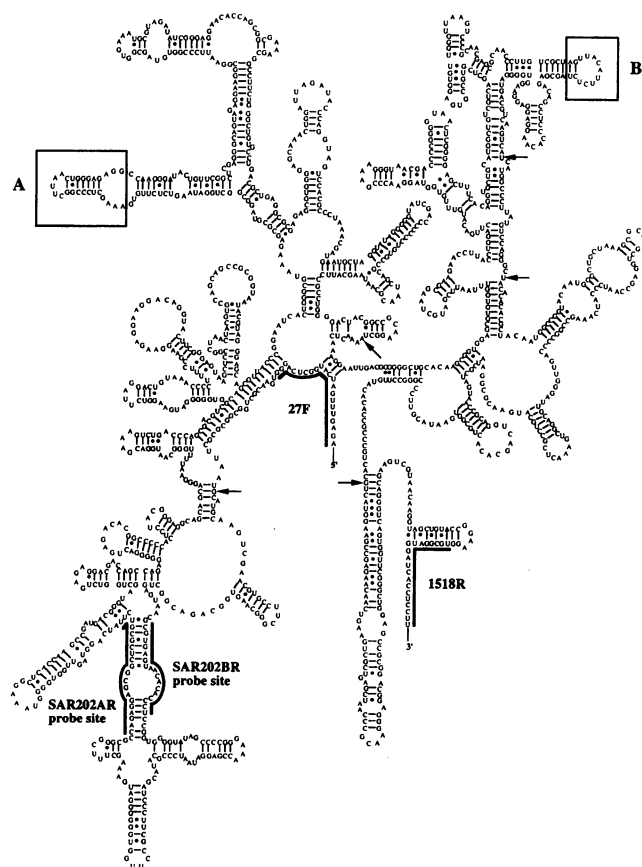


FIG. 2. Secondary structural model of SAR202 16S rRNA. Arrows indicate signature nucleotides of the GNS phylum. The boxed areas indicate unusual secondary structural features that are unique to members of the GNS phylum. (A) Loop structure at *E. coli* positions 607–630. (B) The 15 base pair deletion in the region of positions 1123–1147.

ing. It mismatched the first probe, SAR202AR, at two positions. Therefore, a second probe (SAR202BR), which matched the sequences of both clones SAR202 and SAR307, was synthesized and used in further hybridization experiments with environmental nucleic acids. The specificity of both the SAR202AR and SAR202BR probes was evaluated empirically using blotted arrays of rDNAs from cultured and uncultured bacterioplankton. No examples of crosshybridization to unrelated genes were encountered (data not shown). The program CHECK_PROBE was used to compare the probes to the files of the Ribosomal RNA Database Project, which represent over 2000 prokaryotic species. No 16S rRNA sequences with three or fewer mismatches to the SAR202 probes were found.

The sequencing of additional genes provided evidence that the SAR202 gene cluster is highly diverse. A region of about 400 bases in length between *E. coli* positions 936 and 1346 was chosen for sequencing because it included the unusual deletion between residues 1123 and 1147, which was present in all of the sequenced genes. A phylogenetic tree based on partial sequences is shown in Fig. 3. Two sets of genes were identical, but collectively, the six sequenced genes encompassed substantial variation. The program CHECK_CHIMERA was used to examine the possibility that the genes discussed in this study might be artifacts of *in vitro* recombination events. CHECK_CHIMERA revealed no evidence of multiple evolutionary origins for the genes. As with other gene clusters that have been described previously, it is likely that the data we present underestimate the actual genetic diversity of the group.

Herein, we use the term cluster to refer to a monophyletic set of novel genes that are more similar to each other than to

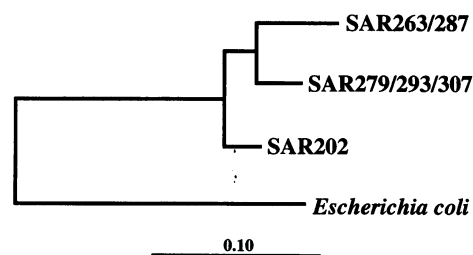


FIG. 3. Phylogenetic tree showing relationships among six rDNA clones that hybridized to the SAR202R probe. The tree was inferred by the maximum likelihood method from 395 nucleotide positions. The root of the tree was determined using the *E. coli* 16S rDNA gene. The scale bar indicates substitutions per nucleotide position.

any rRNA gene from a known organism. When multiple rDNA genes are sequenced from natural systems, such sets of closely related genes are often encountered. The potential explanations for this phenomenon are fourfold. (i) The sequence variation is artifactual, (ii) the variation is due to paralogous evolution within a gene family, (iii) sequence divergence has occurred within large populations that are in essence one “species,” or (iv) the different genes correspond to related lineages of microbes each having a unique relationship to the environment. Although this phenomenon is not yet well understood, a close examination of data such as compensatory base changes across helices has made it clear that artifacts do not explain most of this variation, although they may account for some small fraction of it (26). Such artifacts might result from chimeric genes, sequencing errors, and errors resulting from the misincorporation of bases by polymerase (27). Furthermore, it is known that the cultivable taxa of marine *Synechococcus* Group A and *Prochlorococcus marinus* are phenotypically and ecologically unique but are related by sequence similarities greater than 0.95 (28). The term “phylo-type” also conveniently lumps together related environmental 16S rRNA sequences for the purpose of discussion (8).

The distribution of the SAR202 gene cluster in the water column of the Atlantic and Pacific Oceans was investigated using probes SAR202AR and SAR202BR (Fig. 4). Both probes yielded similar results; however, the hybridization experiments performed with the SAR202AR probe were performed at low stringency and are not presented here. In these experiments, DNA samples from seven depths between the surface and 250 m were amplified with primers for the domain Bacteria. The resulting products were hybridized sequentially to three probes: the two SAR202 probes and the 338R probe, which is specific for the domain Bacteria. In both oceans, the distribution of the SAR202 gene cluster reached a distinct maximum at the lower boundary of the chlorophyll maximum. The depths of the DCM were quite different for the two profiles. In the more oligotrophic Atlantic samples, the DCM was located at 125 m, whereas in Pacific samples, which were collected over the continental slope, the DCM was located at 55 m. The results indicated that bacteria containing the SAR202 gene cluster are potentially significant members of open ocean bacterioplankton communities and further suggested that they are adapted to growth in a discrete zone of the water column.

The phylogenetic group encompassed by the SAR202B probe accounted for about 2.5% of amplicons at the maximum values in vertical profiles. Both the distribution of genes in the clone library from 250 m and the distribution of SAR202 genes in mixtures of amplicons prepared from vertical profiles (Fig. 4) depend on the relative efficiencies of gene amplification during the PCR. Although the fidelity of PCR has been the subject of quantitative studies that addressed the amplification of mixtures of 16S rRNA genes, the potential for biases introduced in such experiments by PCR is not well understood

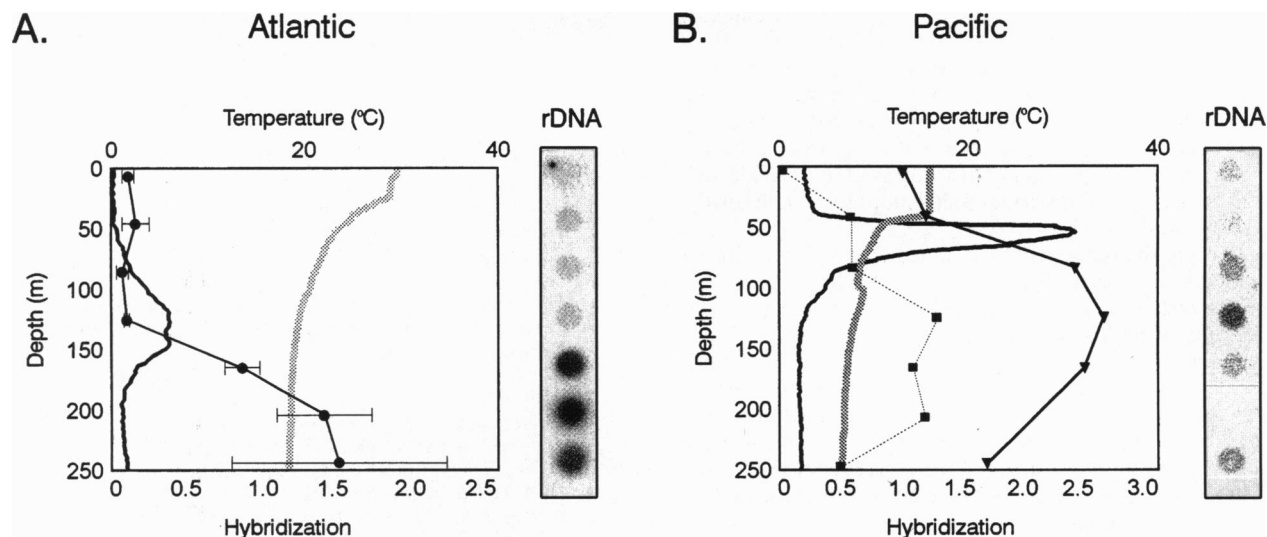


FIG. 4. The distribution of the SAR202 cluster in HMW RNA and 16S rRNA amplicons prepared from plankton samples. Hybridization to HMW RNA is expressed in relative units; hybridization to 16S rDNA amplicons is expressed in absolute units as percent. The SAR202BR probe was used in these experiments, but the SAR202AR probe gave similar results. (A) Samples collected from the upper 250 m at BATS during a stratified period in August 1991. The means and standard deviations are shown for triplicate PCR reactions from a single nucleic acid sample (●). (B) Plankton samples collected in October 1993 at a station located on the continental shelf (ca. 800-m water depth) off the Oregon coast. The values for a single hybridization to rDNA amplicons (▼) and values for hybridization to HMW rRNA from the same sample (■) both indicated that the peak in the SAR202 distribution was located beneath the DCM. Temperature (shaded lines) and chlorophyll a fluorescence (dark lines) are also shown. There was no DNA sample from 200 m for the blot from the Oregon coast.

(29). Other biases could be introduced by the selective recovery of DNA during sampling or the selection of specific template genes due to variations in the structure of hybridization sites. The PCR primer sites used in this study are nearly universally conserved within the Bacteria domain. Thus, major biases due to sequence variation at the target site would not be predicted from the available data, although much remains unknown about microbial diversity.

To provide further support for a stratified location of SAR202 in the water column, probe SAR202BR was hybridized to 10 vertical profiles of RNA from BATS (0–250 m) and 30 consecutive monthly RNA time-series samples from 0 and 200 m. The high, low, and average for the vertical profiles is shown in Fig. 5, along with similar data for *Prochlorococcus* from the same HMW RNA blot. The *Prochlorococcus* hybridization data demonstrate a vertical distribution for these organisms that is very similar to that obtained independently by flow cytometry (30). We show it to demonstrate that hybridization results are consistent with other types of data and also to put the peak in the SAR202 distribution in perspective; prochlorophytes are obligate phototrophs and are most abundant in the lower region of the euphotic zone. Hybridization values for the SAR202 probes were highest in the lower region of the mixed layer and appeared to decline in samples collected below 200 m. Moreover, strong support for the hypothesis that SAR202 was more abundant at 200 m than at 0 m was provided by hybridization to the 30 consecutive time series samples. A one-tailed *t* test assuming unequal variances indicated that the SAR202 cluster was 2.5 times more abundant at 200 m than at 0 m ($P = 2.33 \times 10^{-5}$). The cellular abundance of bacterioplankton, as measured by direct counting methods, varies with depth and sample, and averages ca. 20% less at 200 m than 0 m at BATS. As with other uncultured species, the absolute abundance of SAR202 cannot be accurately estimated from the hybridization of oligonucleotides to RNA because no pure SAR202 RNA is available for the standardization of hybridization results.

In addition to their significance for oceanic ecology, the organisms detected in this study also add new information to the debate over the possibility that the earliest bacteria were

thermophiles. Experimental support for a thermophilic origin of life mainly comes from phylogenetic studies of 16S rRNA genes (31). In particular, these arguments focus on taxa that have evolutionary origins close to the node connecting the three domains of life. Gogarten and colleagues (32) and Iwabe and colleagues (33) established the approximate location of the root of the universal tree using duplicated genes. Although the exact position of the root is unknown, it is presumed to be in the vicinity of the trifurcation. The 16S rDNA clones we have found in the Sargasso Sea cannot be from thermophilic

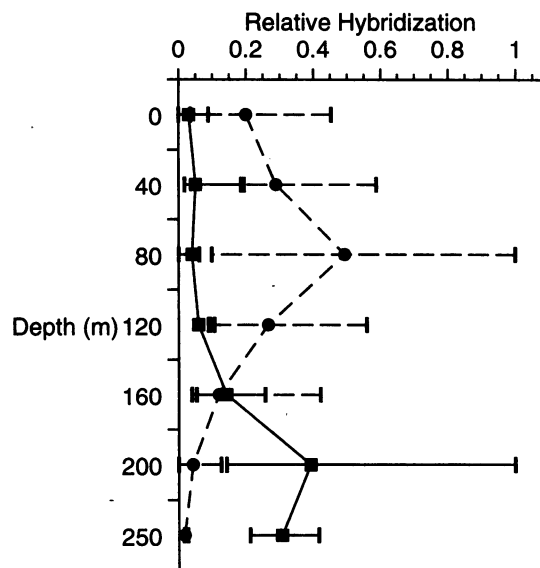


FIG. 5. Average and range of SAR202BR hybridization to picoplankton HMW RNA from 10 depth profiles collected between August 1991 and July 1993 in the Western Sargasso Sea (■). Also shown, for reference, is the distribution of *Prochlorococcus*, as measured by hybridization of the *Prochlorococcus* probe to the same HMW RNA blots (●). Flow cytometry has shown that the abundance of *Prochlorococcus* sp. in the Sargasso Sea is greatest in the vicinity of the DCM, as confirmed here by probe hybridization.

organisms since no thermophilic habitats are reasonably nearby. Instead, it seems that SAR202, the earliest branch of the GNS phylum yet found, represents a mesophilic organism. This implies that the common ancestor of this phylum may not have been a thermophile. It is interesting to note that the group I marine Crenarchaeotes, which were also discovered by the sequencing of 16S rRNA genes from seawater, provide evidence that raises questions about a thermophilic origin for the domain Archaea (16, 17).

Determining the biogeochemical roles of novel, uncultured bacterioplankton species such as the SAR202 group is a challenging task. One avenue for addressing such questions is the collection of ecological data, such as those shown in shown in Figs. 4 and 5, that demonstrate that the SAR202 group is ecologically specialized. The stratification of SAR202 in vertical profiles is most likely a consequence of differential growth and mortality of this population since no physical factors have been noted that could account for the concentration of these cells at these depths. The lower surface layer, where the highest abundance of the SAR202 group was encountered, is oxic. For the 10 dates sampled (Fig. 5), the oxygen concentration in the upper 250 m at BATS varied between 195 μM and 234 μM . It seems unlikely to us that this variation in oxygen concentrations would influence microbial metabolism significantly. The phylogenetic relationships of the SAR202 group to other members of the GNS phylum are too distant to support conclusions about metabolic features that might be shared in common by some or all members of the phylum. However, it is interesting to note that, while the peak in the distribution of the SAR202 gene cluster occurs below the photosynthetic compensation depth for most species, sufficient light may be available to support aerobic photo-heterotrophic metabolism such as that found in the distantly related species *Heliothrix oregonensis* (34). Presently, this possibility is neither more nor less plausible than other credible alternatives, for example, that the SAR202 group are aerobic heterotrophs, similar to *Herpetosiphon*. Their actual biogeochemical role in the oceans must remain unknown pending further study.

The stratification of planktonic bacteria in ocean ecosystems has been known previously for oxygenic bacterioplankton (cyanobacteria and prochlorophytes), but this is the first report showing that significant species from other phylogenetic groups may also have stratified distributions in the surface layer. Additional data from our laboratory indicates that 16S rRNA genes from some other uncultured bacterioplankton also are highly stratified (35). Thus, we argue that the findings we present here provide support for a general model of vertically structured bacterioplankton populations in the surface layer. This model may influence perceptions of bacterioplankton ecology. A microbial community that is constant in composition throughout the surface layer might behave quite differently in response to physical mixing than a community that is stratified, since in the former case one need only consider changes in water chemistry to gauge the impact of mixing events, whereas in the latter case, the impact of redistribution on the physiological activity of the species would become an important consideration.

Note. Additional, very diverse members of the SAR202 gene cluster have been encountered during further random sequencing of the same clone library described in this manuscript.

We are grateful to the Bermuda Biological Station BATS group and the crew of the research vessel *Weatherbird II* for support. In particular, we thank Anthony Michaels for his excellent critical advice, Ann Close for expert oceanographic technical support, and Doug Gordon for the blot from the Pacific Ocean. The Bermuda Biological Station for

Research contribution number for this publication is 1384. This work was supported by National Science Foundation Grant OCE 9016373 and Department of Energy Grant FG0693ER61697.

- Karl, D. M., Knauer, G. A. & Martin, J. H. (1988) *Nature (London)* **322**, 438–441.
- Azam, F. & Cho, B. C. (1988) *Nature (London)* **322**, 441–443.
- Karl, D. M., Letelier, R., Hebel, D., Tupas, L., Dore, J., Christian, J. & Winn, C. (1995) *Nature (London)* **373**, 230–234.
- Ducklow, H. W. (1983) *BioScience* **33**, 494–501.
- Waterbury, J. B., Watson, S. W., Valois, F. W. & Franks, D. G. (1986) *Can. Bull. Fish. Aquat. Sci.* **214**, 71–120.
- Giovannoni, S. J., Mullins, T. D. & Field, K. G. (1995) in *Molecular Ecology of Aquatic Microbes*, ed. Joint, I. (Springer, New York), pp. 217–248.
- Ward, D. M., Bateson, M. M., Weller, R. & Ruff-Roberts, A. L. (1992) *Adv. Microb. Ecol.* **12**, 219–286.
- Reysenbach, A. N., Wickham, G. S. & Pace, N. R. (1994) *Appl. Environ. Microbiol.* **60**, 2113–2119.
- Olsen, G. J., Lane, D. L., Giovannoni, S. J., Pace, N. R. & Stahl, D. A. (1986) *Annu. Rev. Microbiol.* **40**, 337–366.
- Burggraf, S., Olsen, G. J., Stetter, K. O. & Woese, C. R. (1992) *Syst. Appl. Microbiol.* **15**, 352–356.
- Stahl, D. A., Flesher, B., Mansfield, H. W. & Montgomery, L. (1988) *Appl. Environ. Microbiol.* **54**, 1079–1084.
- Poulsen, L. K., Ballard, G. & Stahl, D. A. (1993) *Appl. Environ. Microbiol.* **59**, 1354–1360.
- Giovannoni, S. J., DeLong, E. F., Olsen, G. J. & Pace, N. R. (1988) *J. Bacteriol.* **170**, 720–726.
- Mullins, T. D., Britschgi, T. B., Krest, R. L. & Giovannoni, S. J. (1995) *Limnol. Oceanogr.* **40**, 148–158.
- DeLong, E. F., Franks, D. G. & Alldredge, A. L. (1993) *Limnol. Oceanogr.* **38**, 924–934.
- DeLong, E. F. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5685–5689.
- Fuhrman, J. A., McCallum, K. & Davis, A. A. (1992) *Nature (London)* **356**, 148–149.
- Fuhrman, J. A., McCallum, K. & Davis, A. A. (1993) *Appl. Environ. Microbiol.* **59**, 1294–1302.
- Michaels, A. F., Knap, A. H., Dow, R. L., Gundersen, K., Johnson, R. J., Sorenson, J., Close, A., Knauer, G. A., Lohrenz, S. E., Asper, V. A. & Tuel, M. (1994) *Deep-Sea Res.* **41**, 1013–1038.
- Oyaizu, H., Debrunner-Vossbrink, B., Mandelico, I., Studier, J. A. & Woese, C. R. (1987) *Syst. Appl. Microbiol.* **9**, 47–53.
- Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271.
- Sanger, F., Nicklen, S. & Coulsen, R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Larsen, N., Olsen, G. J., Maidak, B. L., McCaughey, M. J., Overbeek, R. T., Macke, J., Marsh, T. L. & Woese, C. R. (1993) *Nucleic Acids Res.* **21**, 3021–3023.
- Felsenstein, J. (1985) *Evolution* **39**, 783–791.
- Sgaramella, V. & Khorana, H. G. (1972) *J. Mol. Biol.* **72**, 427–444.
- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. (1990) *Nature (London)* **345**, 60–63.
- Shuldiner, A. R., Nirula, A. & Roth, J. (1989) *Nucleic Acids Res.* **17**, 4409.
- Urbach, E., Robertson, D. & Chisholm, S. W. (1992) *Nature (London)* **355**, 267–270.
- Suzuki, M. & Giovannoni, S. J. (1996) *Appl. Environ. Microbiol.* **62**, 625–630.
- Olson, R. J., Chisholm, S. W., Zettler, E. R., Altabet, M. A. & Dusenberry, J. A. (1990) *Deep-Sea Res.* **37**, 1033–1051.
- Achenbach-Richter, L., Gupta, R., Stetter, K. O. & Woese, C. R. (1992) *Syst. Appl. Microbiol.* **9**, 34–39.
- Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6661–6665.
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9355–9359.
- Pierson, B. K., Giovannoni, S. J., Stahl, D. L. & Castenholz, R. W. (1985) *Arch. Microbiol.* **142**, 164–167.
- Gordon, D. A. & Giovannoni, S. J. (1996) *Appl. Environ. Microbiol.* **62**, 1171–1177.