Microbial Ecology

Phylogenetic Screening of Ribosomal RNA Gene-Containing Clones in Bacterial Artificial Chromosome (BAC) Libraries from Different Depths in Monterey Bay

M.T. Suzuki[1], C.M. Preston[2], O. Béjà[2], J.R. de la Torre[2], G.F. Steward[2] and E.F. DeLong[2]

(1) Chesapeake Biological Laboratory, University of Maryland Center for Environmental Sciences, Solomons, MD 20688, USA
(2) Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA

Received: 29 September 2003 / Accepted: 26 February 2004 / Online publication: 9 November 2004

Abstract

Marine picoplankton are central mediators of many oceanic biogeochemical processes, but much of their biology and ecology remains ill defined. One approach to better defining these environmentally significant microbes involves the acquisition of genomic data that can provide information about genome content, metabolic capabilities, and population variability in picoplankton assemblages. Previously, we constructed and phylogenetically screened a Bacterial Artificial Chromosome (BAC) library from surface water picoplankton of Monterey Bay. To further describe niche partitioning, metabolic variability, and population structure in coastal picoplankton populations, we constructed and compared several picoplankton BAC libraries recovered from different depths in Monterey Bay. To facilitate library screening, a rapid technique was developed (ITS-LH-PCR) to identify and quantify ribosomal RNA (rRNA) gene-containing BAC clones in BAC libraries. The approach exploited natural length variations in the internal transcribed spacer (ITS) located between SSU and LSU rRNA genes, as well as the presence and location of tRNA-alanine coding genes within the ITS. The correspondence between ITS-LH-PCR fragment sizes and 16S rRNA gene phylogenies facilitated rapid identification of rRNA genes in BAC clones without requiring direct DNA sequencing. Using this approach, 35 phylogenetic groups (previously identified by cultivation or PCR-based rRNA gene surveys) were detected and quantified among the BAC clones. Since the probability of recovering chimeric rRNA gene sequences in large insert BAC clones was low, we used these sequences to identify potentially chimeric sequences from previous PCR amplified clones deposited in public databases. Full-length SSU rRNA gene sequences from picoplankton BAC libraries, cultivated bacterioplankton, and nonchimeric RNA genes were then used to refine phylogenetic analyses of planktonic marine gamma *Proteobacteria*, *Roseobacter*, and *Rhodospirillales* species.

Introduction

Genomic libraries of large DNA fragments derived from mixed microbial communities provide useful access to the genomes of naturally occurring microorganisms [7, 40, 45]. These resources have a wide variety of applications, including genomic walking from phylogenetic markers to genomically characterize indigenous microbes [4, 6, 7, 35, 42, 45], biochemical analyses of heterologously expressed proteins [4, 43, 44], and microbial population genetic studies [5, 42]. In the marine environment, Bacterial Artificial Chromosome (BAC) libraries constructed from picoplankton have revealed the prevalence of a new photoprotein (proteorhodopsin) in several widespread yet uncultivated bacterial groups [4, 12, 41] and led to the description of marine bacterial genes involved in anoxygenic photosynthesis [6]. Similar approaches have been used to describe rRNA gene-containing fragments from uncultivated *Archaea* [35] and *Acidobacterium* [30] groups in

Present address of O. Béjà: Department of Biology, Technion-Israel Institute of Technology, Haifa 32000, Israel
Present address of J.R. de la Torre: Department of Microbiology, University of Washington, Seattle, WA 98195, USA
Present address of G.F. Steward: Department of Oceanography, University of Hawaii, Honolulu, HI 96822, USA
Present address of E.F. DeLong: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
(M.T. Suzuki and C.M. Preston) These authors contributed equally to this paper.
Correspondence to: E.F. DeLong; E-mail: delong@mit.edu

soil samples. Large insert (BAC) libraries have also been screened for specific genes [6, 12] and enzymatic activity [9, 22, 32]. Finally, BAC termini have been sequenced to estimate the phylogenetic and metabolic representation of genes cloned in BAC libraries ([35], DeLong et al., unpublished data).

In order to describe the phylogenetic diversity within BAC libraries, most environmental genomic studies have depended on direct rRNA gene surveys of these libraries and sequencing of rDNA-containing BAC clones, or alternatively on surveys PCR-amplified rRNA of a parallel sample. Screening of rRNA gene-containing clones in BAC libraries has previously relied on multiplex PCR amplification of the SSU rRNA genes [30, 40] or the SSU rRNA-ITS-LSU rRNA region [7]. However, detection of rRNA gene-containing BAC clones by PCR is complicated by the low copy number of BACs in *Escerichia coli*, and the presence of contaminating *E. coli* chromosomal DNA. Previous attempts to minimize the effects of contaminating *E. coli* DNA have included predigesting purified plasmids with a DNAse specific for linear DNA [7, 30] or inhibition of *E. coli* DNA amplification using terminator primers [30, 40]. Further screening based on SSU rRNA genes has included restriction fragment length polymorphism (RFLP) analysis to identify false positives and rRNA gene amplification cloning and sequencing from BAC clone pools [30, 40]. In contrast, amplification across the internal transcribed spacer region of the rRNA gene uses the natural length heterogeneity of this region [17, 27, 39, 46] to distinguish BAC clone rRNA genes from contaminating *E. coli* amplicons [7]. However, none of the above approaches are particularly suitable for high-throughput screening of multiple BAC libraries.

We previously reported the construction and analysis of a large insert BAC library from Monterey Bay surface bacterioplankton [7]. This BAC library (EBAC or EB000) was shown to contain several groups of uncultivated *Bacteria* and *Archaea* [7]. Since several bacterioplankton groups have depth-specific distributions [23, 47], three additional large insert libraries were constructed from bacterioplankton collected at 80 m, 100 m, and 750 m depths in the Monterey Bay. The phylogenetic identification and quantification of rRNA gene-containing BAC clones in each of the four libraries was determined using a novel, high-throughput PCR analysis: internal transcribed spacer, length heterogeneity PCR (ITS-LH-PCR). The ITS-LH-PCR method was based on the measurement of naturally occurring length heterogeneity of the ITS region, as well as the presence and the location of the tRNA-alanine gene within the ITS. The new screening approach facilitated rapid enumeration and phylogenetic identification of rRNA gene-containing BAC clones in the libraries, and helped to refine analyses of phylogenetic relationships among naturally occurring picoplankton.

Methods

DNA Sampling and BAC Library Construction. The construction of a BAC library from a surface seawater sample in Monterey Bay was previously described [7]. Seawater from 80 m depth was collected on 23 July 1999 at station M1 (36°45.50N 122°02.10W) in the Monterey Bay from multiple casts using a rosette of Niskin bottles aboard the research vessel (RV) *Point Lobos*. Approximately 1000 L was collected, prefiltered through a GFA filter (Millipore, Billerica, MA), and concentrated by tangential flow filtration with a model DC-10L system using a 30,000-Da cutoff hollow fiber cartridge (H10P30-20, Amicon) to 500 mL final volume.

Seawater from the oxygen minimum layer was collected on 11 April 2000, at a station in the Monterey Bay (36°41.1319N 122°02.3727W), using the PISUS (Picoplankton In Situ Underwater Sampler), mounted on the remotely operated vehicle (ROV) *Ventana*. The oxygen minimum layer (750 m, $[O_2] = 0.27$ mL^{-1}) was located using a Sea-Bird O_2 sensor (Sea-Bird Electronics, Inc., Bellevue, WA). Prior to deployment, PISUS hoses, carboys, and submersible pump were filled with sterile, deionized water. At 750 m depth, seawater drawn into PISUS was prescreened through a 30-μm Nitex mesh and concentrated by tangential flow filtration with Pellicon 2 system (Millipore) equipped with a 0.22-μm filter (P2GVPPV20, Millipore) using a submersible pump driven by the ROV hydraulics. The sampler was run for 3.7 h under the following conditions: 45–48 psi inlet pressure, 0–3 psi outlet pressure, 4–5 L min^{-1} flow rate, and back pressure at 34 L min^{-1}. Approximately 1110 L was concentrated *in situ* by the PISUS to 12.4 L final volume. Upon return to the surface, the sample was further concentrated by tangential flow filtration with an Amicon CH2 hollow fiber filtration system to 280 mL final volume.

Cells in both 80-m and 750-m concentrates were pelleted at 30,000 *g* at 4°C, using a SS34 rotor in a model Sorvall RC26 plus centrifuge (Kendro, Newtown, CT). Pelleted cells were resuspended in 0.2-μm filtered seawater and mixed with an equal volume of molten, 40°C, 1% Seaplaque GTG (80 m sample) or InCert (750 m sample) agarose (Cambrex, East Rutherford, NJ), drawn into a modified 1 mL syringe, and placed in ice to solidfy the agarose. The agarose-embedded cells were lysed and gel-embedded chromosomal DNA extracted as previously described [7, 45].

Large genomic fragments of DNA for BAC cloning were prepared by partial *Hind*III digestion (New England Biolabs, Beverly, MA) of DNA-containing agarose slices and size-fractionated by pulsed field gel electrophoresis (PFGE) as previously described [7]. Optimal restriction enzyme concentrations were determined empirically for each agarose plug. Gel regions containing genomic DNA

in the 150–300 and 300–400 kilobasepair (kbp) regions were excised. Gel slices were dialyzed three times with 1× GELase buffer, melted at 65°C for 10 min, cooled to 45°C, and digested with 1 U of GELase (Epicentre Technologies, Madison, WI) per 100 mg of gel for 1 h at 45°C, followed by enzyme inactivation at 70°C for 10 min. Ligation of insert DNA into the pIndigoBAC536 vector and transformation into DH10B electrocompetent cells was performed as previously described [7].

Average insert size in each library was estimated by measuring the sizes of ≥20 linearized BACs per library. BAC DNA was purified by alkaline lysis [2], and 5 μL was digested for 2 h with 0.2 U of the restriction endonuclease *Not*I (Promega, Madison, WI). Linearized BACs were discriminated by PFGE on a CHEF DR II system (Bio-Rad) using the following conditions: 1% Seaplaque GTG agarose (Cambrex) gel in 0.5× TBE (89 mM Tris borate, 2 mM EDTA, pH 8.0) buffer at 6 V cm^{-1}, 5–15 s pulse time, for 13 h at 12°C [7]. The gel was stained with 0.5 μg mL^{-1} of ethidium bromide, destained in water, and scanned using a FluorImager fluorescence imager (Amersham, Piscataway, NJ).

DNA Sampling and Fosmid Library Construction. Seawater from 100 m depth was collected on 21 February 2002 at station M1 (36°45.50N 122°02.10W) in the Monterey Bay from multiple casts using a CTD rosette aboard the RV *Point Lobos*. One thousand four hundred L was collected, prefiltered through a GFA filter (Millipore, Billerica, MA), and concentrated by tangential flow filtration with a model DC-10L system using a 30,000-Da cutoff hollow fiber cartridge (H10P30-20, Millipore) to 330 mL final volume. The hollow fiber cartridge was subsequently backflushed with 8 L of filtrate and concentrated to 500 mL. Primary and secondary concentrates were further concentrated by tangential flow filtration using a Pellicon XL50 system with a 30-kDa NMWL Cartridge (Pellicon 2 Maxi Filter, Millipore) to 15 mL final volume. Cells in these concentrates were pelleted at 12,000 × *g* in a model 5415D microcentrifuge (Brinkmann, Westbury, NY). The primary pellet was embedded into Seaplaque GTG agarose (Cambrex) as described above. The secondary pellet was frozen at –80°C until it was processed.

Because of the difficulties related to obtaining sufficient purified DNA from agar plugs to be used for BAC cloning, DNA from the secondary pellet was extracted and used instead for the construction of large insert libraries using a fosmid vector. The pellet was resuspended in sucrose lysis buffer (40 mM EDTA, 50 mM Tris HCl, pH 8.0, 0.75 M sucrose) containing 0.5 mg mL^{-1} proteinase K (Fisher, Fairlawn, NJ) and 1% SDS (Sigma, St Louis, MO) and incubated at 55°C for 20 min. The cell lysate was incubated at 70°C for 5 min and nucleic acids were extracted twice using phenol:chloroform:IAA (25:24:1, Sigma) and once with chloroform:IAA (24:1, Sigma). Crude nucleic acids were purified using a Centricon 100 (Millipore) spin filter according to the manufacturer's instructions. DNA was further purified by CsCl buoyant equilibrium centrifugation, as previously described [48].

Environmental DNA was cloned using the EpiFOS Fosmid Library Production Kit (Epicentre) following the manufacturer's protocol. Briefly, purified DNA was end repaired according to the manufacturer instructions and size-fractionated by pulsed field gel electrophoresis (PFGE) on a CHEF-DR-II system (Bio-Rad, Hercules, CA) using a 1% SeaPlaque GTG agarose (Cambrex) under the following conditions: 12°C, 6 V cm^{-1} for 16 h and 20–40 s pulse time in 1× TAE (40 mM Tris acetate, 1 mM EDTA, pH 8.0) buffer. The gel was subsequently stained with SYBR Gold (Molecular Probes, Eugene, OR) and viewed on a Dark Reader transilluminator (Clare Chemical Research, Dolores, CO). Gel regions containing genomic DNA in 35–50 kbp regions were excised. The end-repaired and size-selected DNA from gel slices was recovered by gelase treatment as described earlier but without dialysis, concentrated and washed using TE buffer on a Centricon 30, DNA was ligated into the pEpiFOS5 vector, packaged *in vitro* using MaxPlax packaging extracts, and transduced into *E. coli* EPI100 according to the manufacturer's instructions (Epicentre).

Library Screening: Multiplex PCR. Multiplex PCR was performed on BAC and fosmid DNA purified from pooled 96-well microtiter dishes [7]. Briefly, purified BAC and fosmid DNA was digested overnight with Plasmid-Safe DNAse (Epicentre) to remove linear *E. coli* chromosomal DNA. Fifteen plates from the 80-m and 750-m large insert libraries were initially screened for the presence of rRNA gene containing clones using PCR as previously described [7]. Three different primer combinations were used to maximize the identification of different bacterial rRNA operons resulting from a single plate: (l) SSU27F (AGAGTTTGATCCTGGCTCAG) [13] and LSU1933R [1]; (2) SSU27F and BactLSU66R (CACGTCTTTCATCGSCT); and (3) SSU1074F (ATGGCTGTCGTCAGCTCGTG) and BactLSU66R. Primers ArchSSU20F [33] and ArchSSU958R [13] were used to screen for plates with clones containing 16S rRNA genes from *Archaea*. PCR reactions (20 uL) contained 1× *Taq*-Plus Precision buffer (Stratagene, La Jolla, CA), 0.25 μM dNTPs (Promega), 0.1 μM forward and reverse primer, 0.05 U/μL *Taq*Plus Precision DNA polymerase, and 1 μL plasmid-safe treated BAC DNA. Reactions were carried out in an AB9700 thermal cycler (Applied Biosystems, Foster City, CA) under the following conditions: initial denaturation at 93°C for 3 min, 30 cycles of 93°C for 30 s, 55°C for 30 s, and 72°C for 1 min 30 s, followed by a final extension at 72°C for 7 min. PCR products were run in 1×

modified TAE (40 mM Tris acetate, 0.1 mM EDTA, pH 8.0) in a 1% agarose gel (Fisher). Multiplex PCR amplicons with different sizes were excised from the gel, and purified with UltraFreeDA spin columns (Millipore) according to the manufacturer's instructions, and ethanol precipitated. Purified PCR products were either sequenced directly or cloned into the pCR2.1 vector using the Original TA Cloning Kit (Invitrogen, Carlsbad, CA). Ribosomal RNA genes were sequenced by dideoxynucleotide termination using Big Dye Chemistry v3.0 and 320 nM of the primers used for PCR in a AB3100 genetic analyzer (Applied Biosystems). All sequences were trimmed for vector sequences using the software Sequencher (Genecodes Co, Ann Arbor, MI), and aligned using the software ARB [31]. Alignments were then manually inspected and corrected. The phylogenetic affiliation of the clones was determined by adding sequences to a tree containing approximately 19,000 total sequences (tree version: tree_all_sep97) using the ARB_PARSIMONY tool and the POS_ VAR_ BY_PARSIMONY filter.

Screening for rRNA Genes with ITS-LH-PCR. Plasmid-safe treated DNA from pooled plates was used as the template for ITS-LH-PCR reactions. The forward primer BactSSU1406F (TGYACACACCGC CCGT) was fluorescently labeled at the 5′ end with either 6-FAM or HEX (Proligo, Boulder, CO). Two different primer sets were used in separate reactions for each plate pool: FAM-labeled BactSSU1406F and BactLSU66R (CACGTCTTTCATCGSCT) to amplify the entire ITS plus flanking SSU and LSU rRNA gene regions (referred to hereafter as the ITS fragment), and HEX-labeled BactSSU1406F and tRNAalaR (TGCAAGKCAGG TGCTCT) for the fragment between positions homologous to position 1406 of the SSU rRNA gene of *E. coli* [10] and the tRNA-alanine gene (referred to hereafter as the tRNA fragment). In a final volume of 10 μL, PCR reactions contained 1× Platinum *Taq* buffer (Invitrogen), 200 μM dNTP, 3 mM $MgCl_2$, 0.5 μM forward and reverse primers, 0.025U/μL Platinum *Taq* DNA polymerase (Invitrogen), and 1 μL of plasmid-safe treated BAC DNA from pooled plates. Reactions were run under the following conditions on a GeneAmp 9700 (Applied Biosystems): initial denaturation and enzyme activation step at 94°C for 2 min followed by 15 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. One μL of each PCR reaction was combined with 9 μL of 1:0.03 formamide:GS2500 Size Standard (Applied Biosystems, Foster City, CA), denatured at 94°C for 2 min, and separated by capillary electrophoresis using an Applied Biosystems 3100 Genetic Analyzer, equipped with 36- or 80-cm capillaries to discriminate the labeled fragments. Sizes of the fragments were estimated using the Genescan software (Applied Biosystems) and the GS2500 size standard (Applied Biosystems).

The phylogenetic identity of each fragment pair (or single FAM-labeled fragment when the tRNA alanine gene was absent) on each plate was then determined by comparison to reference fragment sizes of previously sequenced rRNA genes [20, 39, 46]. To determine the phylogenetic identity of unidentifiable fragment pairs, PCR reactions with the above primers were performed, products resolved on a 3% NuSieve agarose (Cambrex) gel run for 2 h 70 V in modified 1× TAE (40 mM Tris acetate, 100 μM Na_2EDTA pH 8.0), and the unknown band excised from the gel and sequenced as described above. The SSU rRNA genes from several BAC clones identified using ITS-LH-PCR were fully sequenced as described above, except that additional primers (519R [29], 1100R (AGGGTTGCGCTCGTTG), 1406F [29], and PROK1541R [48]) were also used. All sequences were aligned using the software ARB [31] followed by manual inspection and adjustment of the alignment, and their affiliation determined by their placement into a tree as described above. In addition, fragment-pair sizes from previously published marine bacterioplankton rRNA gene sequences that included the ITS region were also estimated and used as a reference.

Phylogeny of Planktonic Gamma Proteobacteria, Roseobacter Group, and Rhodospirillales. We used SSU rRNA gene sequence in BAC clones in order to refine the phylogeny of typical marine bacterioplankton groups containing cultivated representatives, or exclusively consisting of environmental rRNA gene sequences retrieved by PCR cloning (and therefore subject to chimera formation). Preliminary phylogenetic trees based on full-length, nonchimeric sequences (only cultured organisms and genomic library clones) were constructed for the gamma *Proteobacteria*, the *Roseobacter* group, and the *Rhodospirillales.* Sequence alignments were constructed using the ARB_EDIT software and hypervariable regions with questionable homology (*E. coli* positions 183–193, 204–213, 840–846, and 1134–1139 for the gamma *Proteobacteria*; *E. coli* positions 69–100 and 1003–1036 for the *Roseobacter* group; and *E. coli* positions 71–97, 197–217, 841–846, 1004–1010, and 1025–1036 for *Rhodospirillales*) were excluded from further analysis using a manually edited filter (filter 1). A second filter was created using the ''filter by base frequency'' tool in ARB that excluded positions with ambiguous characters and positions in the alignment where gaps were more frequent than characters (filter 2). Phylogenetic reconstruction was based on the remaining positions after both filters were applied (1269 positions for 53 sequences in the gamma *Proteobacteria*, 1204 positions for 90 sequences in the *Roseobacter* group, and 1112 positions for 61 sequences in the *Rhodospirillales*). The sequences were exported and phylogenetic analyses performed by neighbor-joining using the PHYLIP package [15]. Next we searched

Table 1. Origin and properties of coastal picoplankton BAC libraries

	Library			
Parameter	*EB000*	*EB080*	*EF100*	*EB750*
Collection date	17 Mar 99	23 Jul 99	21 Feb 02	11 Apr 00
Cloning approach	BAC	BAC	FOSMID	BAC
Cells for DNA extraction	2.5×10^{11}	1.5×10^{11}	1.4×10^{11}	5×10^{10}
Clones in library	8352	12,672	14,976	1536
Average insert size (kbp)	80	74	40	60
Genomes	223	313	200	31
Clones screened for rDNA	2400	8928	14,976	1536
rRNA genes detected (%)	3.0	3.3	0.9	1.8

sequences with putative chimeric origin among the remaining full-length (>1300 bp) published plankton rRNA gene-containing clones belonging to the gamma *Proteobacteria*, the *Roseobacter* group, and the *Rhodospirillales*, by constructing trees using either the 5′ or the 3′ end of the rRNA gene, and searching for sequences with different placement according to the 5′ or 3′-end phylogeny. The sequences (87 for gamma *Proteobacteria*; 23 for the *Roseobacter* group; and 21 for the *Rhodospirillales*) were added to the preliminary non chimeric trees using ARB_PARSIMONY with the same filters as above, except that two versions of filter 2 were created. The first version contained positions lower than the *E. coli* position 756; the second contained positions higher than the *E. coli* position 757. Putatively chimeric sequences were excluded from further analysis. SSU rRNA gene sequences from cultured strains, genomic library clones, and nonchimeric planktonic PCR clones were exported to PHYLIP format using filters 1 and 2 and phylogenetic trees for the gamma *Proteobacteria*; the *Roseobacter* group and the *Rhodospirillales* were constructed by neighbor-joining using the PHYLIP package. Finally, bootstrap analyses (100 replicates) were performed using the PHYLIP package.

Results

Library Descriptions. Four large insert libraries, three BAC-based and one fosmid-based, were constructed from bacterioplankton collected at four different depths and dates in the Monterey Bay [7, this study]. These four libraries in total comprised 37,536 BAC clones (Table 1). Average insert sizes ranged from 40 kbp for the 100 m fosmid library to 80 kbp for the surface BAC library (Table 1). Assuming a 3.0 megabasepair (Mbp) average genome size, we recovered 767 genome equivalents, ranging from 31 genome equivalents in the EB750 library to 313 genome equivalents in the EB080 library (Table 1). The percentage of rRNA gene-containing clones revealed by all screening methods combined ranged from 0.9% for the EF100 library to 3.3% for the EB080 library (Table 1).

Phylogeny of Gamma Proteobacteria, Roseobacter Group, and the Rhodospiralles. Since it is unlikely that rRNA genes in BAC clones represent chimeras, and since several gamma *Proteobacteria* genes on BACs were affiliated with previously undescribed clades, we evaluated the phylogeny of the gamma *Proteobacteria*, as well as the *Roseobacter* and *Rhodospiralles* groups. Fifteen full-length SSU rRNA gene sequences belonging to the gamma *Proteobacteria*, three full-length SSU rRNA gene sequences belonging to the *Roseobacter* group, and six full-length SSU clones from the *Rhodospiralles* group were determined from BAC clones. These sequences, along with those from cultivated organisms (boldface taxa in Figs. 1, 2, and 3), were used to create a phylogenetic tree of nonchimeric sequences. Onto this tree, we added full sequences (>1300) from previously published bacterioplankton rRNA gene sequences by ARB_PARSIMONY, using 5′-end and 3′-end comparisons to identify putative chimeras. Among 21 *Rhodospirillales*, 22 *Roseobacter* groups, and 66 gamma *Proteobacteria* clones added, we detected zero, one, and 11 putative chimeras, respectively.

The resulting gamma *Proteobacteria* tree was used to define eight previously unnamed groups (clades underlined in Fig. 1), most of which were represented in the BAC libraries. Clades ARCTIC95B-7 and SAR86-I were exceptions and might be represented by chimeric sequences, since all clones in each of the clades were recovered by PCR and originated from a single transformation [3, 46]. Novel clades were named after the first published full-length (>1300) rRNA gene sequence belonging to each clade. Two exceptions are the SAR156 and the AGG47 clades that were so named for historical reasons. The SAR156 name has been previously used interchangeably to represent the SAR86 clade, since in earlier versions of the ribosomal database project (RDP) (i.e., release 7.0, July 1998), sequences belonging to the SAR86 group [38, 46] were included in the "environmental clone SAR156" group [47]. The SSU rRNA gene sequence for clone SAR156 (GenBank accession number L35469) is partial (815 bp), and therefore was not included in our phylogenetic reconstruction. We added the

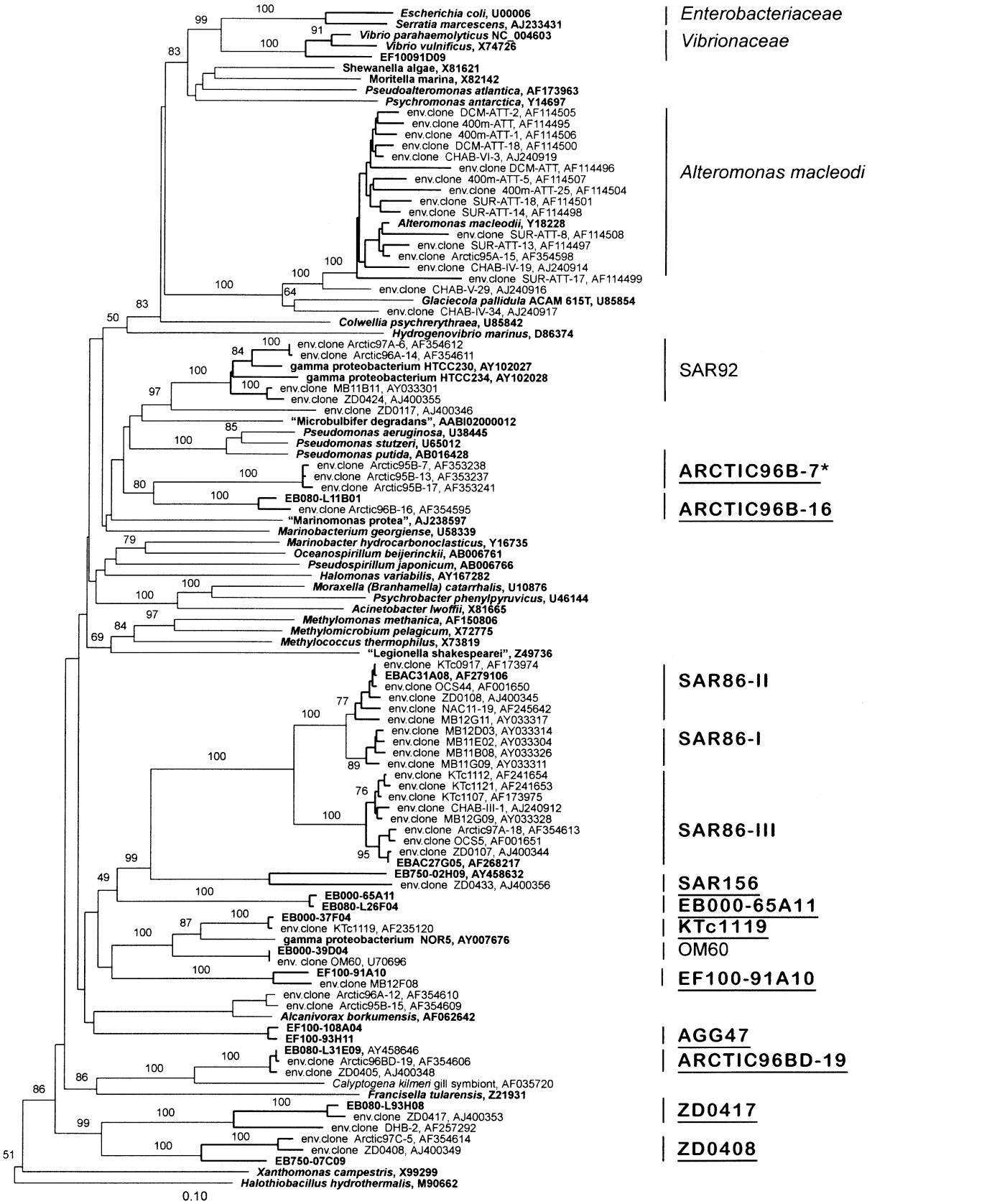

Figure 1. Phylogenetic reconstruction (neighbor joining) of the planktonic gamma *Proteobacteria* including near-full-length SSU rRNA sequences (>1400 base pairs) of isolates and BAC clones (boldface), and nonchimeric environmental rRNA clones (prefix env. clone). *Burkholderia cepacia* (not shown) was used as the outgroup. Bootstrap values were based on 100 replicated trees. * denotes groups that contain only rRNA gene clones retrieved in PCR-based clone libraries. Underlined clades represent newly defined clades. Bolded clades do not contain cultured representatives.

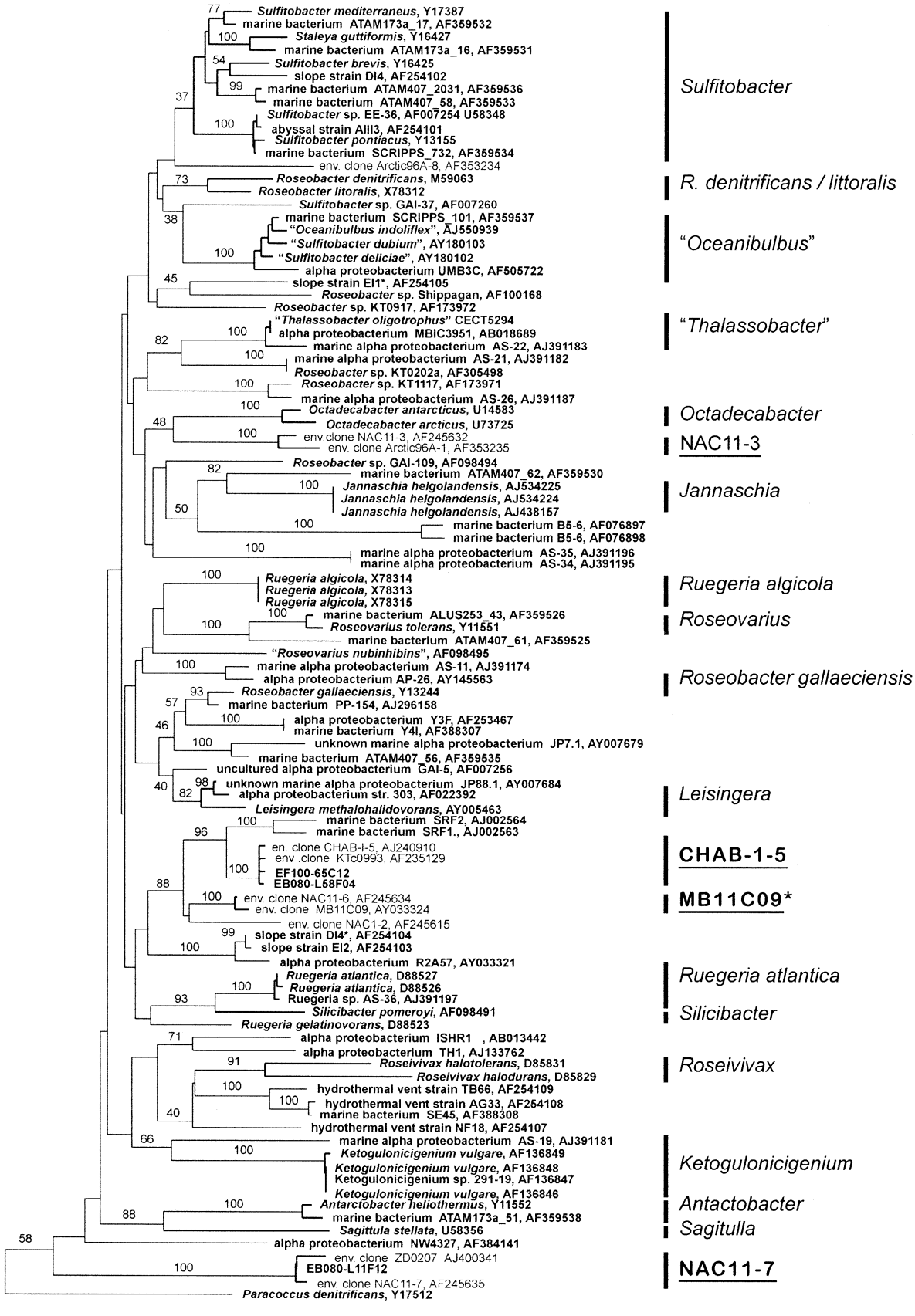

Figure 2. Phylogenetic reconstruction (neighbor joining) of the *Roseobacter* group including near-full-length SSU rRNA sequences (>1400 base pairs) of isolates and BAC clones (boldface), and nonchimeric environmental rRNA clones (prefix env. clone). Bootstrap values were based on 100 replicated trees. * denotes groups that contain only rRNA gene clones retrieved in PCR-based clone libraries. Underlined clades represent newly defined clades. Bolded clades do not contain cultured representatives.

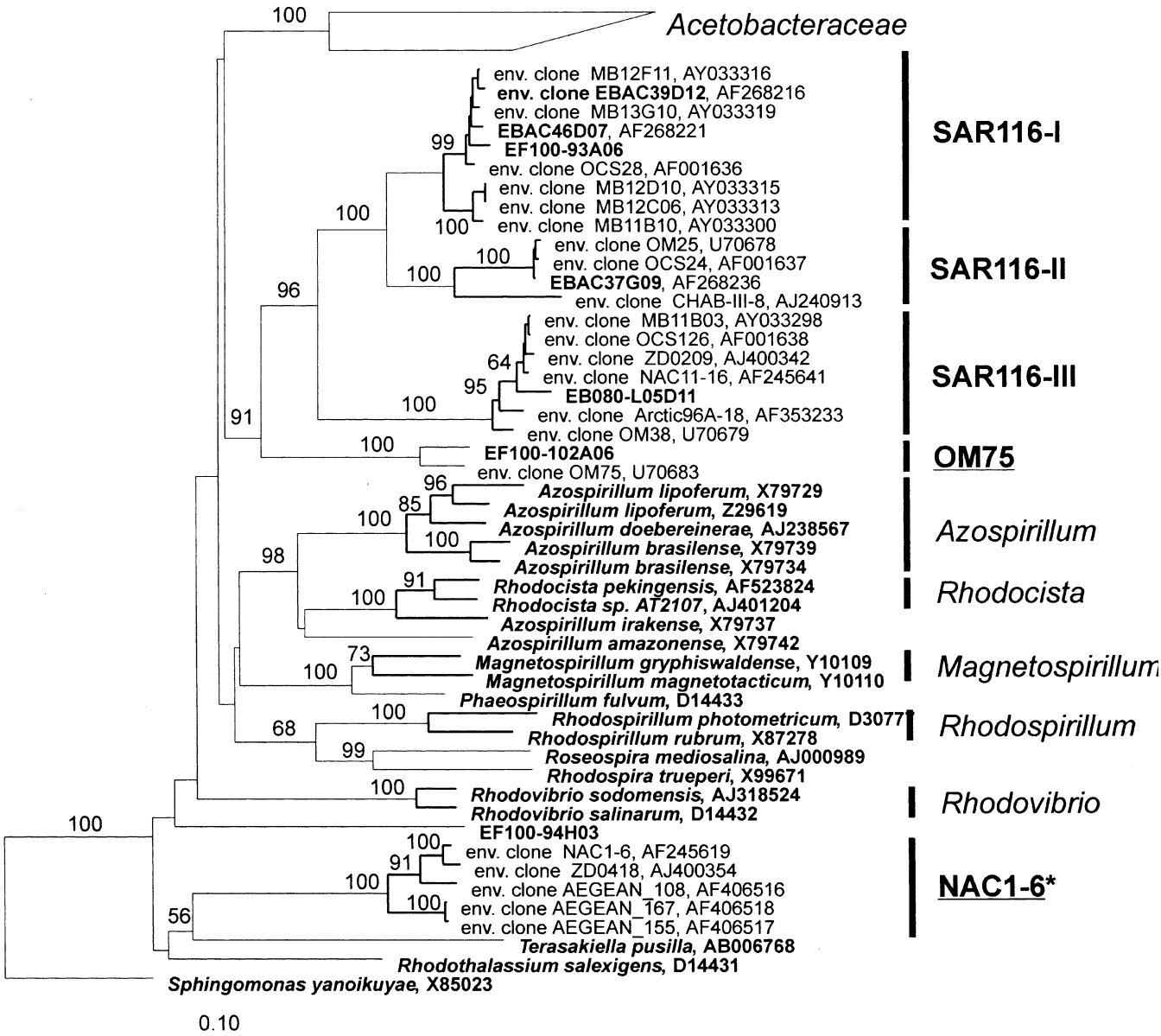

Figure 3. Phylogenetic reconstruction (neighbor joining) of the class *Rhodospirillales* including near-full-length SSU rRNA sequences (>1300 base pairs) of isolates and BAC clones (boldface), and nonchimeric environmental rRNA clones (prefix env. clone). Bootstrap values were based on 100 replicated trees. * denotes groups that contain only rRNA gene clones retrieved in PCR-based clone libraries. Underlined clades represent newly defined clades. Bolded clades do not contain cultured representatives.

SAR156 rRNA gene sequence by ARB_PARSIMONY to the tree represented in Fig. 1, using the same masks described, and it clearly clusters with BAC clone EB750-02H09 and SSU rRNA gene clone ZD0403 (data not shown). A DNA distance matrix further supported this affiliation (data not shown). The clade descriptor "SAR156" is used to distinguish the clade containing EB750-02H09 and ZD0403, from the SAR86 clade proper.

The phylogeny of the *Roseobacter* group points to the difficulty of using SSU rRNA genes to define relationships among members of this group. Although very closely related sequences, particularly those belonging to a single species, formed clades with high bootstrap support, the relationships between these clades were less clear. Several previously named genera (i.e., *Roseobacter*, *Ruegeria*, and *Silicibacter*) appear to be polyphyletic according to our phylogenetic reconstruction. This phylogenetic reconstruction was also used to define three previously unnamed clades that consisted exclusively of not-yet-cultured organisms (clades underlined in Fig. 2), most of which contained sequences from BAC clones. Novel clades were named after the first published full-length (>1300) rRNA gene sequence belonging to each

clade of the *Roseobacter* group. Clade MB11C09 contained neither a cultivated representative or BAC clone and thus may consist of chimeric sequences, although this seems unlikely since clones in this clade originate from different libraries [25, 46]. The clade NACII-3 in Fig. 2 contains a sequence from an isolate recently recovered via high throughput culturing efforts (HTTC152 [11]). The SSU rRNA gene sequence in GenBank for this isolate (AY102029) was partial (647 bp), and thus it was not included in our analysis. The affiliation of HTC152 was determined by ARB_PARSIMONY using the same masks described above and was found to be >99% similar to members of the NAC11-3 clade.

The phylogeny of the *Rhodospirillales* is in good agreement with previous classification schemes (e.g. "Taxonomy Outline of the Prokaryotic Genera" [21], and the Ribosomal Database Project–II preview classification). Sequences for most genera in both the *Acetobacteraceae* (not shown) and the *Rhodospirillaceae* (Fig. 3) were placed in monophyletic clades with high bootstrap support. A notable exception is the genus *Azospirillum*, which appears to be polyphyletic. Furthermore, previously defined subgroups in the SAR116 clade [46] were strongly supported by the placement of BAC clones in each of the three subgroups (Fig. 3). This phylogenetic reconstruction also defined two previously unnamed clades that consisted exclusively of not-yet-cultured organisms (OM75 and NAC1-6), named after the first published full-length (>1300) rRNA gene sequence belonging to the clades. Clade OM75 contained a BAC sequence from the 100 m fosmid library, whereas clade NAC1-6 was composed exclusively of SSU rRNA gene clones recovered by PCR and could therefore be represented by sequences of chimeric origin. However, this is very unlikely since several clones in the NAC1-6 clade originated from different clone libraries [25, 50]. Finally, BAC clone EF100-94H03 was not closely related to any of the clades in the *Rhodospirillales* and might represent another clade in this group.

ITS-LH-PCR. ITS-LH-PCR is a high-resolution method using capillary electrophoresis that relies on the natural ITS length heterogeneity, as well as the presence and location of the tRNA alanine gene within the ITS. We found a good correspondence between paired ITS and tRNA fragment sizes (Table 2), with phylogenetic identity as determined by SSU rRNA phylogenies (Figs. 1–3). This congruence was determined either from previously published rRNA operon sequences or from BAC clones sequenced in this study (Table 2). ITS fragment sizes ranged from 366 bp for the marine *Actinobacteria* to 1228 bp for *Ruegeria algicola*. Sizes for the tRNA fragment ranged from 296 bp for the OM75 clade to 651 bases for *Ruegeria algicola*. Genes coding the tRNA-alanine were absent in many groups including *Pseudoalteromonadales*, SAR86, SAR156, EB000-65A11, SAR116, and the marine *Actinobacteria*. Several groups show overlapping ITS or tRNA fragment size ranges (i.e., ITS fragments for *Pelagibacter* and the SAR116 clade), but when combined, the sizes of both fragments allowed putative identification without the need for sequencing. Furthermore, we were able to distinguish the three subclades of the SAR86 clade [46] based on ITS-LH-PCR peak sizes (Table 2).

Figure 4 shows an electropherogram of labeled PCR fragments amplified from a pooled plate and the putative identity of each of the peaks or peak pairs. Despite plasmid-safe DNAse treatment of purified BAC clones, chromosomal DNA from *E. coli* was still detectable, with four ITS peaks (654 bases, 645 bases, 638 bases, and 563 bases), and one tRNAala peak (378 bp). These peaks however were easily identified and removed from analysis (Fig. 4). Furthermore, *E. coli* amplified fragments could be used as positive controls of ITS-LH-PCR reaction in particular for plates with no rRNA gene containing clones. Finally, in some cases we observed tRNA fragments without a corresponding ITS fragment, indicating that extra degenerate positions might be necessary to improve the performance of the BactLSU66R.

Phylogenetic Composition of BAC Libraries. Initially, a subset of each library was screened with three different primer sets specific to *Bacteria*, to maximize fragment discrimination in agarose gels. Primer pair SSU-1074F/BactLSU66R produced the best discrimination and the amount of SSU rDNA sequence obtained (~400 bp per fragment) allowed phylogenetic placement using ARB_PARSIMONY. We detected many phylogenetic groups typically found in marine plankton including *Pelagibacter* (SAR11), SAR86, SAR156, *Roseobacter* group, *Synechoccocus* group, SAR116, marine *Actinobacteria*, marine delta *Proteobacteria* (SAR324), OM60, and ARCTIC96BD-19 in the BAC libraries (Table 3). However, because of the relatively poor resolution of bands in agarose gels and the large amount; of sequencing required to resolve uncertainties, the ITS-LH-PCR method was developed to screen BAC libraries for bacterial rRNA operons with higher resolution and throughput.

We used ITS-LH-PCR to analyze the previously described surface library [7] (EB000) as well as the 80 m (EB080), 100 m (EF100) and 750 m (EB750) BAC libraries. The number of rRNA gene containing clones in each 96-well microtiter dish ranged between zero and eight. As expected, BAC libraries with larger average inserts sizes contained a higher proportion of rRNA genes in comparison to the fosmid library. The percentage of rRNA gene-containing clones was 3.0%, 3.3%, 0.9%, and 1.8% in the EB000, EB080, EF100, and EB750 libraries, respectively (Table 1). Based on the sizes of ITS-LH-PCR fragment pairs, we assigned clone affiliation at the group

Table 2. ITS-LH-PCR fragment sizes predicted based on sequence information and measured for typical picoplanktonic groups

	Size (bp)					Size (bp)			
	Predicted		Measured			Predicted		Measured	
Phylogenetic clade	*ITS fragment*	*tRNA fragment*	*ITS fragment*	*tRNA fragment*	*Phylogenetic clade*	*ITS fragment*	*tRNA fragment*	*ITS fragment*	*TRNA fragment*
GAMMA PROTEOBACTERIA					BETA PROTEOBACTERIA				
Vibrionaceae ($n = 1$)	–	–	584	no	OM43 ($n = 2$)	768–771	373	774	375
SAR92 ($n = 1$)	691	390	690–692	391–393	*Nitrosocbccus* ($n = 1$)	–	–	490–492	no
Pseudomonas ($n = 1$)	–	–	727	no	*Nitrosomonas* ($n = 1$)	883	394	887	394
Pseudoalteromonas ($n = 1$)	555	no	–	–	DELTA PROTEOBACTERIA				
ARCTIC96B-16 ($n = 1$)	1098	577	–	–	Marine delta proteobacteria SAR324 ($n = 1$)	700	431	–	–
Marinomonas ($n = 1$)	–	–	559–561	no					
SAR86 ($n = 8$)[a]	400–477	no	401–481	no	FBC				
SAR86-I ($n = 4$)	400–404	no	401–406	no	MB1IE04	657	417	657	419
SAR86-II ($n = 2$)	439–463	no	440–465	no	EB080-L05	688	466	–	–
SAR86-III ($n = 2$)	468–477	no	469–481	no	*Flexibacter johnsonii*	–	–	806	519
SAR156 ($n = 3$)	483–498	no	498–499	no	EB080-L08E11	–	–	806	521
EB000-65A11 ($n = 1$)	696	no	–	–	strain R2A103	–	–	526	no
KTc119 ($n = 1$)	944	483	–	–	SAR406 (*Fibrobacter*)				
OM60 ($n = 1$)	898	381	–	–	EB750-03B02	531	366		–
EF100-9IA10			1057	480	MB13C05	593	355		–
Alcanivorax ($n = 1$)	–	–	391–392	no					
AGG47	–	–	670–673	no	*Vermcomicrobiales* ($n = 2$)	671	365	670–672	367–368
ARCTIC96BD-19 ($n = 4$)	781	373	785–799	379–381					
ZD0417 ($n = 1$)	704	435	–	–	*Cyanobacteria*[b]				
ZD0408 ($n = 1$)	686	426	–	–	Low B/A *Prochlorococcus* Clade I ($n = 4$)	760–763	398–399	764	400
ALPHA PROTEOBACTERIA					Low B/S *Prochlorococcus* Clade II ($n = 18$)	750–759	398–406	–	–
Roseobacter Group					High B/A *Prochlorococcus* Clade I ($n = 3$)	845–846	440	–	–
Ruegeria algicola ($n = 1$)	1228	651	–	–	High B/A *Prochlorococcus* Clade II ($n = 5$)	879	450	–	–
Antarctobacter ($n = 1$)	1158	405	–	–	High B/A *Prochlorococcus* Clade III ($n = 1$)	906	451	–	–
R. denitrificans/litoralis ($n = 2$)	1050–1054	497	–	–	High B/A *Prochlorococcus* Clade IV ($n = 2$)	1042–1044	455–457	–	–
Roseovarius ($n = 1$)	1030	506	–	–	Marine A *Synechococcus* Clade I ($n = 4$)	971–997	455–457	–	–
NAC11-7 ($n = 2$)	–	–	1127–1131	427–428	Marine A *Synechococcus* Clade II ($n = 5$)	987–988	457	–	–
Strain R2A57	1001	352	1004–1107	339–341	Marine A *Synechococcus* Clade III ($n = 6$)	1017–1023	459–461	–	–
MB11C09 ($n = 1$)	958	433	959–962	433–435	Marine A *Synechococcus* Clade IV ($n = 2$)	987–989	456–457	994	456–458

(Continued)

Table 2. Continued

Phylogenetic clade	Predicted		Measured		Phylogenetic clade	Predicted		Measured	
	ITS fragment	*tRNA fragment*	*ITS fragment*	*tRNA fragment*		*ITS fragment*	*tRNA fragment*	*ITS fragment*	*TRNA fragment*
CHAB-1-5 (n = 3)	–	–	925–928	436–437	Marine A *Synechococcus* Clade VI (n = 4)	960–972	449–458	–	–
Rhodobacter Group (n =1)		–	1001–1004	442–443	Marine A *Synechococcus* Clade V (n = 3)	993–994	457–458	–	–
EF100-94H03	707	426	705	426	Marine B *Synechococcus* sp. WH8010	1000	459	–	–
SAR116 (n = 8)	587–643	no	586–646	no	Marine B *Synechococcus* sp. WH5701	1073	472	–	–
OM75 (n = 1)	810	296	–	–	Plastid (n = 3)	–	–	491–502	no
Pelagibacter (n = 108)	555–648	362–375	591–647	363–371	Marine *Actinobacteria* (n = 4)	366	no	367–368	no
MB13F01 (n = 1)	633	411	631–633	413–414	Firmicutes (*Haloanaerobium* n = 1)	–	–	531	no

[a]Clades according to [45].
[b]Clades according to [40].
–: not determined.

level for all libraries, and a summarized description of the composition of the libraries determined using all screening methods is shown in Table 3. Combined, in all four BAC libraries, we detected 35 different major bacterial phylogenetic clades [23, 47] (Table 1), with the majority of the clones (58.4%) belonging to clades previously only recovered by cultivation-independent techniques.

Clones belonging to the alpha *Proteobacteria* were members of the SAR11 (*Pelagibacter ubique*) and SAR116 clades, the *Roseobacter* group [23, 47], and clades OM75 and affiliated with clone EF100-94H03 described above. The SAR11 clade was present in all libraries and, although only a small number of clones were sequenced, we were able to retrieve clones from different SAR11 subclades [16, 46], including members of subclades IA and IIA [46]. Furthermore, the distribution of SAR11 clones in the libraries agrees with the hypothesis that these clades are differentially distributed in the water column. Clones belonging to clade IA (which includes the cultivated species *Pelagibacter ubique*) were retrieved in EB000 and EB080, while clones belonging to clade IIA, and clones related to SAR203 (accession number U75255) and SAR220 (accession number U75255), were retrieved in EB750. Members of the *Roseobacter* group were most abundant in the surface and 80 libraries, composing a large proportion of these libraries. Subclade NAC11-7 of the *Roseobacter* group was particularly interesting as members of this clade were only recovered by cultivation-independent studies and it represented the most abundant rRNA gene type when all libraries were combined. Finally, subclades of the *Rhodospirillales* showed contrasting depth distributions as members of the SAR116 clade were detected only in the swallow (0, 80, and 100 m) libraries. Members of the OM75 clade were detected only in EF100, whereas clones affiliated with EF100-94H03 were retrieved in EF100 and EB750.

Members of the gamma subdivision of the *Proteobacteria* were detected in all libraries, although we observed differential depth distribution of several groups belonging to this subdivision. The majority of clones belonging to the SAR86 clade were retrieved in the upper water column (surface and 80 m) while no SAR86 clones were retrieved from EF750. Additional uncultivated groups of the gamma *Proteobacteria* included previously unnamed clades Arctic96B-16, SAR156, EB000-65A11, Ktc1119, EF100-91A10, AGG47, Arctic96BD-19, ZD0417 and ZD0408 (Table 3, Fig. 1). Among these, clade ARCTIC96BD-19 was particularly interesting, since it was the most abundant gamma *Proteobacteria* group and was present in all libraries (Table 3). Previously, clones in the ARCTIC96BD-19 clade were retrieved mostly from marine plankton and were closely related to sulfur-oxidizing symbionts of vesicomyid clams (Fig. 1). Gamma *Proteobacteria* BAC clones closely related to previously

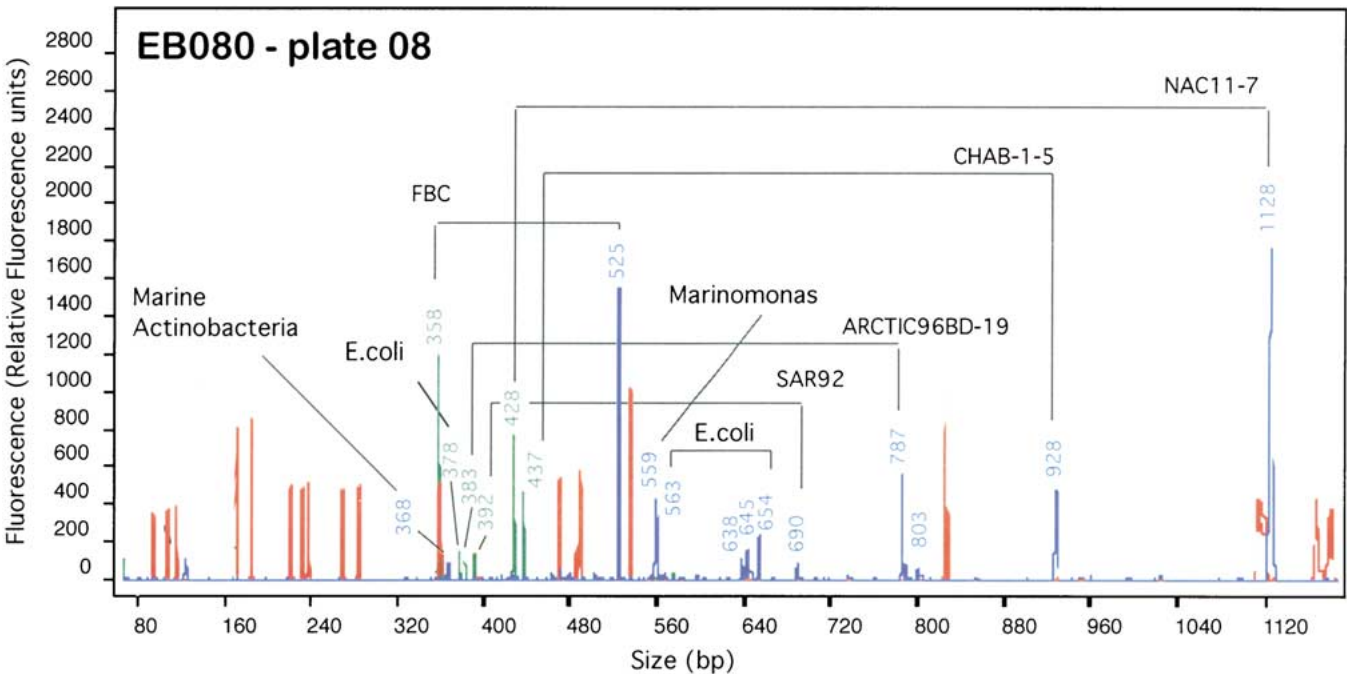

Figure 4. ITS-LH-PCR: Electropherogram of labeled fragments amplified from pooled plate 08 from EB080 run on a 36-cm capillary, showing the ITS fragment peaks in blue and the tRNA fragment peaks in green. The identification of rRNA gene containing clones was based on a comparison to the reference clones listed in Table 2.

cultivated organisms belonged to the genera *Vibrio*, *Pseudoalteromonas*, and *Marinomonas* as well as the SAR92 [11], NOR4 [14], and OM60 [11] clades (Table 3).

Additional previously uncultivated groups represented in the BAC libraries included the marine delta *Proteobacteria* (SAR324 clade [49]) which was particularly abundant in EB750; the *Fibrobacter* phylum (SAR406 clade [26]) present in EB750, the *Verrucomicrobiales*, the marine *Actinobacteria* [37], and the *Haloanaerobium* clade. Also, the *Flavobacterium*, *Bacteroides*, and *Cytophaga* phylum was represented mostly by clones falling into groups composed exclusively of sequences retrieved in cultivation-independent studies, including a group of surface sequences (EBAC43, AF268230; EBAC391, AF2682232 and EBAC40, AF268229) closely related to rRNA gene clone FL7 (L10937). In addition, one surface sequence (EBAC322, AF268231) and one sequence from EB080 (EB080-L08X11) were closely related to rRNA gene clone ARCTIC97A-17 (AF354617). The remaining clones belonged to groups with previously cultivated members including the *Cyanobacteria* genus *Synechococcus*, the delta *Proteobacteria* genus *Nitrospina*, and the beta *Proteobacteria* genus *Nitrosomonas* and OM43 clade [11]. A number of unique fragments pairs remain unidentified, particularly in the EF100 library (47.7%).

Large insert clones containing rRNA genes from the domain *Archaea* were only recovered in the surface and 100 m libraries. A BAC clone containing the LSU rRNA gene from a marine *Euryarchaeota* was previously reported in the surface library and has been completely sequenced [7]. SSU rRNA genes from the marine *Crenarchaeota* (Group I) and marine *Euryarchaeota* (Group II) were detected in equal numbers in the 100 m library, each representing 3.9% of rRNA gene containing clones.

Discussion

In order to access information about the genomic organization and metabolic potential of typical marine picoplankton, we successfully constructed three large insert libraries using DNA isolated from Monterey Bay. Although average insert sizes for fosmid clones are considerably shorter, we were able to retrieve a greater number of clones via transduction in the fosmid libraries compared to BAC libraries using smaller amounts of starting biomass. The smaller required biomass, more rigorous lysis protocol, more random DNA shearing of inserts, and simpler overall cloning procedure confer distinct advantages on the fosmid cloning procedure, relative to the BAC approach [35; Delong et al, unpublished data].

The ITS-LH-PCR approach provided a minimum estimate of the total rRNA gene diversity of the libraries. Complicating factors include possible primer biases, the presence in the same microtiter plates of multiple clones with overlapping sizes, disruption of the ITS region due

Table 3. Percentages of phylogenetic clades in large insert libraries determined by all screening methods

	Library			
Phylogenetic clade	*EB000*	*EB080*	*EF100*	*EB750*
Gamma *Proteobacteria*	16.9	34.5	27.5	22.2
Vibrio	–	–	3.1	–
SAR92	–	1.4	–	–
Pseudoa Iteromonas	–	0.3	–	–
Arctic96B-16	–	1.4	–	–
Marinotnonas	–	0.3	–	–
SAR86-I	2.8	–	–	–
SAR86-II	1.4	2.0	–	–
SAR86III	1.4	3.4	1.6	–
SAR156	–	3.7	–	7.4
EB000-65A11	1.4	0.7	4.7	–
KTc1119	1.4	–	0.8	–
OM60	2.8	0.7	1.6	–
EF100-91A10	–	–	0.8	–
Agg47	–	–	6.3	–
Arctic96BD-19	5.6	19.9	5.5	7.4
ZD0417	–	0.3	–	–
ZD0408	–	–	–	3.7
NOR4	–	–	0.8	–
Methylobacter	–	–	–	3.7
Beta *Proteobacteria*	1.4	3.0	–	–
OM43	1.4	2.7	–	–
Nitrosomonas	–	0.3	–	–
Alpha *Proteobacteria*	50.7	39.2	11.7	40.7
NAC11-7	21.1	23.6	–	–
CHAB-1-5	5.6	5.7	0.8	–
Other *Roseobacter* clades	7.0	6.4	2.3	–
EF100-94H03	–	–	0.8	3.7
SAR116	11.3	1.4	0.8	–
Pelagibacter (SAR11)	5.6	2.0	7.0	37.0
OM75	–	–	2.3	–
Delta *Proteobacteria*	–	0.7	–	7.4
SAR324	–	0.3	–	7.4
Nitrospina	–	0.3	–	–
FBC	8.5	1.4	–	–
SAR406 (Fibrobacter)	–	–	–	3.7
Verrucomicrobiales	–	2.0	4.7	3.7
Cyanobacteria	12.7	0.3	1.6	_
Synechococcus	1.4	0.3	1.6	–
Chloroplast	11.3	–	–	–
Marine *Actinobacteria*	–	3.0	1.6	–
Firmicutes	–	–	0.8	–
Marine *Crenarchaeota*	–	–	3.9	–
Marine *Euryarchaeota*	1.4	–	3.9	–
Unidentified	8.5	15.5	44.5	22.2

Groups in boldface are represented by rRNA genes retrieved exclusively by cultivation independent approaches. –: not detected.

to cloning, and the possibility that some bacteria may not have linked SSU and LSU rRNA genes. Fortunately, the observation that the vast majority of groups previously recovered in picoplankton rRNA gene libraries [23, 47] were also recovered in our genomic libraries indicates that the diversity detected in BAC libraries is not significantly underestimated. Notably absent from our libraries were sequences belonging to the SAR202 cluster related to *Chloroflexus* and *Herpetosiphon* [24]. Members of SAR202 might have unlinked SSU and LSU rRNA genes or mismatches with the tRNAalaR and BactLSU66R primers, or they might simply have been present in low numbers in our samples. The later seems unlikely, since SAR202 has been detected at depths ranging from 50 m to 3000 m [18, 24], with a peak abundance just below the deep chlorophyll maximum [24]. Screening of the libraries with SAR202-specific primers should be used in the future to resolve this issue.

A number of studies have used the length heterogeneity of rRNA gene internal transcribed spacer regions for the analysis of microbial communities in the environment [i.e., 8, 17, 19]. The techniques employed in these studies (RISA or ARISA) used universally conserved primers to produce fingerprints and to estimate overall microbial community diversity. ITS-LH-PCR differs from RISA or ARISA, since it is a library screening method rather than a community fingerprinting method, and more importantly, two fragment sizes are measured in order to establish the phylogenetic origin of clones. Using ITS-LH-PCR we were able to more comprehensively describe the phylogenetic diversity contained in four BAC libraries from the Monterey Bay. The congruence between combined sizes of the ITS and tRNA-alanine fragment sizes and 16S rRNA gene phylogeny allowed putative phylogenetic identification of a large number of peaks without sequencing. ITS-LH-PCR also appears to be promising as a method for screening of rRNA operon clone libraries (M. Suzuki, J. Kan, F. Chen, and S. Evans, in preparation). Finally, based on our observations that the tRNAalaR and BactLSU66R primers did not produce amplification products from certain rRNA gene containing clones, the use of the technique for community fingerprinting is not advised.

The depths sampled by the BAC libraries correspond to different niches in the pelagic ecosystem, and we were able to recover and identify genomic clones originating from several picoplankton groups not detected in the previously reported surface BAC library [7]. The non-chimeric origin of rRNA genes in BAC clones allowed us to refine the phylogenetic reconstructions within the gamma *Proteobacteria* the *Roseobacter* group, and the class *Rhodospirillales*. Based on these reconstructions we also identified several novel clades of bacterioplankton: 10 in the gamma *Proteobacteria*, four in the *Roseobacter group*, and two in the class *Rhodospirillales*. Most of these clades were composed exclusively of sequences recovered by cultivation-independent studies and helped to explain why the majority of clones in our study (58.4%) belonged

to previously uncultivated clades. This percentage is conservative, since all clones in the *Pelagibacter* clade and the *Flavobacteria–Bacteroides–Cytophaga* bacterial division were included as members of previously cultivated groups. In fact, several members of the *Pelagibacter* clade and FBC phylum belonged to subclades composed exclusively of sequences recovered by cultivation-independent studies. These results further emphasize the utility of large insert DNA libraries for accessing the metabolic and ecological properties of typical marine planktonic microbes.

According to Giovannoni and Rappé [23], nine bacterial groups account for about 80% of the marine plankton clones deposited in the Genbank database. Our analysis of BAC libraries showed that these nine dominant clone groups comprised only 42% of rRNA gene containing inserts. These data could simply reflect the ecological situation in Monterey Bay at the time of our sampling. These differences might also be explained by the fact that most rRNA gene libraries from picoplankton have been constructed using PCR-amplified genes from surface water samples. Another alternative explanation might be differential recoveries in BAC versus PCR amplified rRNA gene libraries. Among the novel clades defined by our phylogenetic reconstruction, ARCTIC96BD-19 comprised a large number of rRNA gene containing clones in the BAC libraries, particularly in EB080. This clade is solely composed of sequences retrieved by culture-independent studies, and high percentages of members of this group have not been previously reported. These high percentages of ARCTIC96BD-19 suggest that, although a number of groups such as *Pelagibacter* and SAR86 are ubiquitous in marine picoplankton [23, 47], other novel and uncharacterized groups may also significantly contribute to picoplankton communities in different ecological contexts, for instance in spatially and temporally dynamic coastal regions.

Among rRNA gene-containing clones we observed in coastal near-surface waters, the *Roseobacter* group appeared by far the most abundant, accounting for 26% of all clones. Moreover, a single subclade (NAC11-7) represented about 65% of all clones in the *Roseobacter* group in EB000 and EB080. These high percentages of NAC11-7 rRNA gene clones indicate we may have sampled blooms of these organisms. In addition, this group may contain a higher number of rRNA operons per genome, a hypothesis that could be directly tested by sequencing these clones. Members of this subclade were previously recovered from samples collected during dimethylsulfoniopropionate (DSMP) producing phytoplankton blooms in the North Atlantic Ocean, and their involvement in DMSP utilization has been suggested [25, 50]. In fact, Zubkov and co-workers [50] showed that members of the NAC11-7 clade represented 70% of rRNA gene clones in a library constructed from a bloom-associated subpopulation that was sorted by flow cytometry and represented 28% of total bacterioplankton cell counts. High chlorophyll-*a* concentrations, indicating phytoplankton bloom conditions, are typically observed at the Monterey Bay sites where libraries EB000 and EB080 were prepared. Thus, the high percentages of NAC11-7 clade rRNA gene-containing BAC clones support its association with phytoplankton in different oceanic provinces, although we cannot speculate on the role of this group in DMSP metabolism. Remarkably, to date, there have been no cultivated strains reported that belong to the NAC11-7 clade, nor is there information regarding the specific metabolic capabilities of organisms of this group, despite its abundance. (However, many other *Roseobacter* do have at least one cultivated member.) This serves as a caution against extrapolating metabolic capabilities of populations from the properties of single strains or clades, especially considering that cell abundances are frequently determined only at the group level (e.g., *Roseobacter* group abundance).

A recent study analyzing the sequence similarity of rRNA genes (groups defined at >97% similarity) recovered from marine bacterioplankton suggested that the species richness of marine bacterioplankton was low [28]. Examination of our analyses of rRNA gene sequence diversity within the gamma *Proteobacteria*, *Roseobacter* group, and the *Rhodospirillales*, even using a 97% cutoff, reveals diverse and distinct clusters of closely related but non-identical rRNA gene sequences. Sequences in some of these gene clusters are known to exhibit depth specific distributions [16, 34], as well as high degrees of genomic [5] and functional [34, 39] divergence. Thus, a 97% cutoff in SSU rRNA gene sequence similarity is expected to severely underestimate the genomic and functional diversity, as well important ecological differences, among picoplankton groups identified by ribosomal RNA gene sequence.

Recent high throughput cultivation methods have resulted in the isolation of several of the previously uncultivated clades including *Pelagibacter* (SAR11), *Roseobacter* NACII-3, OM60, OM43 and SAR92 [11, 36]. In the future, we believe that information obtained via combined cultivation efforts and environmental genomics will provide the basis for determining genome structure, population dynamics, metabolic potential, and biogeochemical relevance of several typical picoplankton groups. The rRNA gene-containing BAC clones reported here are a useful tool for characterizing and interrelating the genomic content and phylogenetic relationships of indigenous marine bacterioplankton.

Data Deposition. Sequences reported here have been submitted to GenBank under the following accession numbers: AY627365 – AY627383.

Acknowledgments

We thank the captain and crews of the RV *Point Lobos*/ROV *Ventana* as well as Erich Rienecker, Lynne Christianson, Virginia Rich, and Mark Jeanette for their assistance during sampling. This work was supported by a grant from the David and Lucile Packard Foundation and NSF Microbial Observatory grant 0084211 to E.F.D. and John Heidelberg (TIGR). O. Béjà was supported by a long-term fellowship from the European Molecular Biology Organization and NSF grant OCE0001619 to E.F.D.

References

1. Amann, RI, Ludwig, W, Schleifer, KH (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol Rev 59(1): 143–169
2. Amemiya, CT, Ota, T, Litman, GW (1996) Construction of P1 artificial chromosome (PAC) libraries from lower vertebrates. In: Birren, B, Lal, E (Eds.) Nonmammalian Genomic Analysis: A Practical Guide, Academic Press, New York, pp 223–256
3. Bano, N, Hollibaugh, JT (2002) Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. Appl Environ Microbiol 68(2): 505–518
4. Béjà, O, Aravind, L, Koonin, EV, Suzuki, MT, Hadd, A, Nguyen, LP, Jovanovich, SB, Gates, CM, Feldman, RA, Spudich, JL, Spudich, EN, DeLong, EF (2002) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science 289(5486): 1902–1906
5. Béjà, O, Koonin, EV, Aravind, L, Taylor, LT, Seitz, H, Stein, JL, Bensen, DC, Feldmen, RA, Swanson, RV, DeLong, EF (2002) Comparative genomic analysis of archeal genotypic variants in a single population and in two different oceanic provinces. Appl Environ Microbiol 68(1): 335–345
6. Béjà, O, Suzuki, MT, Heidelberg, JF, Nelson, WC, Preston, CM, Hamada, T, Eisen, JA, Fraser, CM, DeLong, EF (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. Nature 415(6872): 630–633
7. Béjà, O, Suzuki, MT, Koonin, EV, Aravind, L, Hadd, A, Nguyer, LP, Villacorta, R, Amjadi, M, Garrigues, C, Jovanovich, SB, Feldman, RA, DeLong, EF (2000) Construction and analysis or bacterial artifical chromosome libraries from a marine microbial assemblage. Environ Microbiol 2(5): 516–529
8. Borneman, J, Triplett, EW (1997) Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. Appl Environ Microbiol 63(7): 2647–2653
9. Brady, SF, Chao, CJ, Handelsman, J, Clardy, J (2001) Coning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Org Lett 3(13): 1981–1984
10. Brosius, J, Dull, TJ, Sleeter, DD, Noller, HF (1981) Gene organization and primary, structure of a ribosomal RNA operon from *Escherichia coli.* J Mol Biol 148: 107–127
11. Connon, SA, Giovannoni, SJ (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl Environ Microbiol 68(8): 3878–3885
12. de la Torre, JR, Christianson, LM, Beja, O, Suzuki, MT, Karl, DM, Heidelberg, J, DeLong, EF (2003) Proteorhodopsin genes are distributed among divergent marine bacterial taxa. Proc Natl Acad Sci USA 100(22): 12830–12835
13. DeLong, EF (1992) Archaea in coastal marine environments. Proc Natl Acad Sci USA 89(12): 5685–5689
14. Eilers, H, Pernthaler, J, Glockner, FO, Amann, R (2000) Culturability and *in situ* abundance of pelagic bacteria from the North Sea. Appl Environ Microbiol 66(7): 3044–3051
15. Felsenstein, J (1989) PHYLIP—phylogeny inference package (v3.5). Cladistics 5: 164–166
16. 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(1): 63–70
17. Fisher, MM, Triplett, EW (1999) Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl Environ Microbiol 65(10): 4630–4636
18. Fuhrman, JA, Davis, AA (1997) Widespread Archaea and novel Bacteria from the deep sea as shown by 16S rRNA gene sequences. Mar Ecol Prog Ser 150(1–3): 275–285
19. Garcia-Martinez, J, Acinas, SG, Anton, AI, Rodriguez-Valera, F (1999) Use of the 16S–23S ribosomal genes spacer region in studies of prokaryotic diversity. J Microbiol Methods 36(1–2): 55–64
20. 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(7): 935–948
21. Garrity, GM, Winters, M, Seales, DB. (2001) Taxonomic Outline of the Prokaryotic Genera, 2nd ed., Bergeys Manual of Systematic Bacteriology. Springer Verlag, New York
22. Gillespie, DE, Brady, SF, Bettermann, AD, Cianciotto, NP, Liles, MR, Rondon, MR, Clardy, J, Goodman, RM, Handelsman, J (2002) Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68(9): 4301–4306
23. Giovannoni, SJ, Rappé, MS (2000) Evolution, diversity and molecular ecology of marine prokaryotes. In: Kirchman, DL (Ed.) Microbial Ecology of the Oceans, Wiley, New York, pp 47–84
24. Giovannoni, SJ, Rappe, MS, Vergin, KL, Adair, NL (1996) 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. Proc Natl Acad Sci USA 93(15): 7979–7984
25. Gonzalez, JM, Simo, R, Massana, R, Covert, JS, Casamayor, EO, Pedros-Alio, C, Moran, MA (2000) Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. Appl Environ Microbiol 66(10): 4237–4246
26. Gordon, DA, Giovannoni, SJ (1996) Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific oceans. Appl Environ Microbiol 62(4): 1171–1177
27. Gurtler, V, Stanisich, VA (1996) New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. Microbiology 142(1): 3–16
28. Hagstrom, A, 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(7): 3628–3633
29. Lane, DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt, E, Goodfellow, M (Eds.) Nucleic Acid Techniques in Bacterial Systematics, John Wiley & Sons, New York, pp 115–175
30. Liles, MR, Manske, BF, Bintrim, SB, Handelsman, J, Goodman, RM (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Appl Environ Microbiol 69(5): 2684–2691
31. Ludwig, W, Strunk, O, Westram, R, Richter, L, Meier, H, Kumar, Y, Buchner, A, Lai, T, Steppi, S, Jobb, G, Förster, W, Brettske, I, Gerber, S, Ginhart, AW, Gross, O, Grumann, S, Hermann, S, Jost, R, König, A, Liss, T, Lüßmann, R, May, M, Nonhoff, B, Reichel, B,

Strehlow, R, Stamatakis, AP, Stuckmann, N, Vilbig, A, Lenke, M, Ludwig, T, Bode, A, Schleifer, KH (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32: 1363–1371

32. MacNeil, IA, Tiong, CL, Minor, C, August, PR, Grossman, TH, Loiacond, KA, Lynch, BA, Phillips, T, Narula, S, Sundaramoorthi, R, Tyler, A, Aldredge, T, Long, H, Gilman, M, Holt, D, Osburne, MS (2001) Expression and isolation of antimicrobial small molecule from soil DNA libraries. J Mol Microbiol Biotechnol 3(2): 301–308
33. Massana, R, Murray, AE, Preston, CM, DeLong, EF (1997) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. Appl Environ Microbiol 63(1): 50–56
34. Moore, LR, Rocap, G, Chisholm, SW (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. Nature 393(6684): 464–467
35. Quaiser, A, Ochsenreiter, T, Klenk, HP, Kletzin, A, Treusch, AH, Meurer, G, Eck, J, Sensen, CW, Schleper, C (2002) First insight into the genome of an uncultivated crenarchaeote from soil. Environ Microbiol 4(10): 603–611
36. Rappé, MS, Connon, SA, Vergin, KL, Giovannoni, SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418(6898): 630–633
37. Rappé, MS, Gordon, DA, Vergin, KL, Giovannoni, SJ (1999) Phylogeny of actionbacteria small subunit (SSU) rRNA gene clones recovered from marine bacterioplankton. Sys Appl Microbiol 22(1): 106–112
38. Rappé, MS, Kemp, PF, Giovannoni, SJ (1997) Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. Limnol Oceanogr 42(5): 811–826
39. Rocap, G, Diste, DL, Waterbury, JB, Chisholm, SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. Appl Environ Microbio 68(3): 1180–1191
40. Rondon, MR, August, PR, Bettermann, AD, Brady, SF, Grossmann, TH, Lies, MR, Loiacono, KA, Lynch, BA, MacNeil, IA, Minor, C, Tiong, CL, Gilman, M, Osburne, MS, Clardy, J, Handelsman, J, Goodman, RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66(6): 2541–2547
41. Sabehi, G, Massana, R, Bielawski, JP, Rosenberg, M, DeLong, EF, Béjà, O (2003) Novel Proteorhodopsin variants from the Mediterranean and Red Seas. Environ Microbiol 5(10): 842–849
42. Schleper, C, DeLong, EF, Preston, CM, Feldman, RA, Wu, KY, Swanson, RV (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum.* J Bacteriol 180(19): 5003–5009
43. Schleper, C, Swanson, RV, Mathur, EJ, DeLong, EF (1997) Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum.* J Bacteriol 179(24): 7803–7811
44. Stein, JL, Felbeck, H (1993) Kinetic and physical properties of a recombinant RuBisCO_2 from a chemoautotrophic endosymbiont. Mol Mar Biol Biotechnol 2(5): 280–290
45. Stein, JL, Marsh, TL, Wu, KY, Shizuya, H, DeLong, EF (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J Bacteriol 178(3): 591–599
46. Suzuki, MT, Béjà, O, Taylor, LT, DeLong, EF (2001) Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. Environ Microbiol 3(5): 323–331
47. Suzuki, MT, DeLong, EF (2002) Marine prokaryotic diversity. In: Staley, JD, Reysenbach, AL (Eds.) Biodiversity of Microbial Life: Foundation of Earth's Biosphere, Wiley, New York, pp 209–234
48. Suzuki, MT, Taylor, LT, DeLong, EF (2000) Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5′-nuclease assays. Appl Environ Microbiol 66(11): 4605–4614
49. Wright, TD, Vergin, KL, Boyd, PW, Giovannoni, SJ (1997) A novel ô-subdivision proteobacterial lineage from the lower ocean surface layer. Appl Environ Microbiol 63(4): 1441–1448
50. Zubkov, MV, Fuchs, BM, Archer, SD, Kiene, RP, Amann, R, Burkill, PH (2001) Linking the composition of bacterioplankton to rapid turnover of dissolved dimethylsulphonioprocionate in an algal bloom in the North Sea. Environ Microbiol 3(5): 304–311