

# Marine bacterial microdiversity as revealed by internal transcribed spacer analysis

Mark V. Brown<sup>1,2,\*</sup>, Jed A. Fuhrman<sup>1</sup>

<sup>1</sup>Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089-0371, USA

<sup>2</sup>Present address: NASA Astrobiology Institute, University of Hawaii, Honolulu, Hawaii 96822, USA

**ABSTRACT:** A growing body of evidence suggests analysis of 16S rRNA gene sequences provides only a conservative estimate of the actual genetic diversity existing within microbial communities. We examined the less conserved internal transcribed spacer (ITS) region of the ribosomal operon to determine the impact microdiversity may have on our view of marine microbial consortia. Analysis of over 500 ITS sequences and 250 associated 16S rRNA gene sequences from an oceanic time series station in the San Pedro Channel, California, USA, revealed that the community in this region is composed of large numbers of distinct lineages, with more than 1000 lineages estimated from 3 clusters alone (the SAR11 clade, the *Prochlorococcus* low-B/A clade 1, and the *Roseobacter* NAC11-7 clade). Although we found no instances where divergent ITS sequences were associated with identical 16S rRNA gene sequences, the ITS region showed much greater pairwise divergence between clones. By comparison to our 16S rRNA gene–ITS region linked database, we were able to place all ITS sequences into a phylogenetic framework, allowing them to act as an alternative molecular marker with enhanced resolution. Comparison of SAR11 clade ITS sequences with those available in GenBank indicated phylogenetic groupings based not only on depth but also on geography, potentially indicating localized differentiation or adaptation.

**KEY WORDS:** Microbial ecology · Microdiversity · Intergenic transcribed spacer

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

## INTRODUCTION

The ribosomal operon, which in prokaryotes is generally composed of 16S and 23S ribosomal RNA (rRNA) genes separated by an internal transcribed spacer (ITS) region, is part of the slowly evolving bacterial 'core genome', the set of genes present in all bacteria. Several reports have indicated that organismal gene content closely reflects phylogeny, and that incremental diversification of the core genome reflects the diversification of the 'accessory' genome over time (Feil 2004). However, it remains unclear as to what level phylogenetic markers, if any, may accurately correspond to genomic diversification. In several cases (e.g. Rocop et al. 2002, Seurinck et al. 2003, Jaspers & Overmann 2004) genetic clusters defined by ribosomal ITS sequences have been identified as 'ecotypes' (closely related organisms inhabiting different niches) within an environment. Although the ITS region does

play a role in rRNA processing and may encode for up to 2 tRNA genes, it appears for the most part to lack the requirement for structural integrity that constrains mutation in the ribosomal molecules. Consequently, this region displays a great deal of length and sequence variation which may provide high taxonomic resolution and serve as a fast molecular chronometer by which initial genome diversification and evolutionary speciation may be detected (Schloter et al. 2000). In the present study, we examined the ITS sequence of all clones retrieved during a temporal and spatial analysis of a marine time series station in the San Pedro Channel, California, USA. Our aim was to verify the extent of microdiversity commonly observed in environmental surveys that employ 16S rRNA gene sequence analysis (e.g. Acinas et al. 2004) using an alternative, less conserved marker, and examine what effects this finer resolution may have on our current understanding of marine microbial diversity.

\*Email: mbrown@ifa.hawaii.edu

## MATERIALS AND METHODS

**Study site and sample collection.** Coastal marine water was collected as part of the San Pedro Ocean Time Series (SPOTS) monthly sampling at a station situated approximately midway between San Pedro Harbor and Catalina Island in the San Pedro Channel, California, USA (33° 33' N, 118° 24' W). Months and depths analyzed for this study included: October 2000, 5 m; April 2001, 5 and 14 m (corresponding to the chlorophyll *a* maximum); August 2001, 5 m; December 2001, 5 m; February 2002, 43 m (corresponding to the chlorophyll *a* maximum); May 2003, 150, 500, and 890 m. The time series station displays a complex hydrographic regime, with the deeper waters (particularly below 700 m) being almost anoxic, whilst the euphotic zone is alternately influenced by oligotrophic eddies moving inshore and up the coast or by nutrient-rich water up-welled near Point Conception to the north and swept southward in the California Current. Samples were obtained using 20 l Niskin bottles, sequentially filtered through a 142 mm Whatman GF/F filter followed by a 142 mm Durapore 0.22 µm pore size filter using a positive pressure filtration system, and filters were stored at –80°C until processing. Auxiliary data obtained from the water samples used to create clone libraries, including temperatures, salinities, chlorophyll and nutrient data, can be accessed at [www.usc.edu/microbialobservatory](http://www.usc.edu/microbialobservatory).

**Clone library construction and analysis.** Community DNA was extracted from Durapore filters using hot SDS lysis followed by phenol:chloroform:isoamyl-alcohol extractions and ethanol precipitation as described by Fuhrman et al. (1988). Monthly depth and temporal variability of samples were initially screened using Automated rDNA Intergenic Spacer Analysis (ARISA) (Fisher & Triplett 1999). Clone libraries were created from samples that would enable a wide range of ARISA peak identification, i.e. where peaks representing major bacterial groups varied. PCR amplicons encompassing almost full length 16S rRNA genes along with the associated ITS spacer region were amplified using the primers 27f, 5'-AGAGTTTGATC MTGGCTCAG-3' (bacterial-specific 16S rRNA gene) (Lane 1991) and 23Sr, 5'-GGGTTBCCCCATTCRG-3' (bacterial-specific 23S rRNA gene) (Fisher & Triplett 1999). The 100 µl reaction mixture contained a final concentration of 1× PCR AmpliTaq Gold<sup>®</sup> buffer (Applied Biosystems), 3.5 mM MgCl<sub>2</sub> (Applied Biosystems), 350 µM of each dNTP (Promega), 800 nM of each primer, 40 ng µl<sup>-1</sup> BSA, 5 U of the high fidelity DNA polymerase AmpliTaq Gold<sup>®</sup>, and 0.1 ng µl<sup>-1</sup> of template DNA. The reaction mixture was held at 94°C for 10 min followed by 24 cycles of amplification at 94°C for 40 s, 55°C for 40 s and 72°C for 3 min with a

final step at 72°C for 7 min. Amplification products were purified using a Qiagen MinElute PCR purification column, quantified using Pico Green fluorescence (Molecular Probes) in a Bio-Rad VersaFluor<sup>™</sup> fluorometer and ligated into pGEM<sup>®</sup>-T Easy (Promega) plasmid vectors. Ligation reactions were transformed into JM109 High Efficiency Competent Cells (Promega) and colorimetric screening was used to determine clones containing recombinant vectors. Plasmid DNA was isolated from 3 ml overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen).

Plasmid DNA was sequenced with the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences) and 5 pmol of sequencing primer. Sequencing reactions were prepared and cleaned according to the manufacturer's instructions. Electrophoresis was carried out on an ABI 377XL automated sequencer. Initial sequencing was done with the universal 16S rRNA gene primer 1392F (5'-GYACACACCGCCCGT-3'), to encompass most of the ITS region. Further sequencing was done on multiple representative clones with primers M13F (5'-GTAAAACGACGGCCAGT-3'), M13R (5'-CAGGAAACAGCTATGAC-3'), universal 536F (5'-CAGCMGCCGCGGTAATWC-3') and one of 1520r a (5'-AAGGAGGTGATCCAGCC-3'), b (5'-TAGGAGGTGATCCAGCC-3'), c (5'-AAGGAGGTAATCCAGCC-3'), d (5'-AAGGAGGTGATCCAACC-3'), e (5'-AAGGAGGTGTTCCAGCC-3'), f (5'-AAGGAGATGTTCCAGCC-3'), g (5'-AAAGAGATATTCAGCC-3'), h (5'-AAGGAGGTATTCCAGCC-3'), i (5'-ATGG AGGTGATCCAGCC-3'), j (5'-AAGGAGGTGATCCATCC-3'), k (5'-AAGGAGATAATCCAGCC-3'), l (5'-TAGGAGGTGATCCATCC-3') to provide full length 16S rRNA gene sequences. To conserve resources, not all 16S rRNA gene sequences were completed. Where 2 or more identical ITS sequences were found, at least 2 of the associated 16S rRNA gene regions were sequenced. As the objective of the survey was to examine ITS sequence diversity, 16S rRNA gene sequences were used to ground that diversity in a phylogenetic framework. All raw sequence data were verified manually from electropherograms using the program Chromas version 1.45 ([www.mb.mahidol.ac.th/pub/chromas/chromas.htm](http://www.mb.mahidol.ac.th/pub/chromas/chromas.htm)). Completed clone sequences were checked using the chimera detection programs Chimera Check (<http://rdp8.cme.msu.edu/cgis/chimera.cgi?su=SSU>) and Bellerophon (Huber et al. 2004). Sequences have been deposited in GenBank under the accession numbers DQ009080 to DQ009478. 16S rRNA gene sequences were imported into the Hugenholtz alignment (Hugenholtz 2002) in ARB (Ludwig et al. 2004) aligned to their closest relatives using the ARB alignment tool and manually corrected. 16S rRNA gene phylogenetic trees were generated in ARB using the

neighbour-joining function with appropriate taxon based (e.g. alpha-Proteobacteria) filters, which utilize only conserved bases within each group. Only full length sequences (those covering the range of the filter) were used in the analysis. ITS sequences were aligned in BioEdit ([www.mbio.ncsu.edu/BioEdit/bioedit.html](http://www.mbio.ncsu.edu/BioEdit/bioedit.html)) initially using Clustal W followed by manual editing. Trees inferring ITS sequence relationships were generated in the PHYLIP software package (Felsenstein 2004) using DNAdist and Neighbour programs. As only closely related ITS sequences could be aligned, and as many ITS sequences have relevant insertions or deletions that are not necessarily conserved within even closely related groups, all base pairs were used in the analysis. We were not as interested in the depth or length of branches as the cluster patterns they produced. Bootstrap values were calculated for 100 replicate trees using the programs Seqboot, DNAdist, Neighbour, and Consense from the PHYLIP package. Comparisons of phylogenetic reconstructions using 16S rRNA gene sequences and the ITS region provided a further test for chimeric sequences.

Richness estimates were calculated using the Chao1 richness estimator (Hughes et al. 2001) with 100% sequence identity.

## RESULTS

The ITS sequence was obtained for all 512 clones examined, along with 248 associated 16S rRNA gene sequences (either full length or >1000 base pairs), resulting in the compilation of an extensive ITS sequence database coupled to 16S rRNA gene phylogenetic information. In our database, we did not observe any variation in ITS length or sequence associated with a single 16S rRNA gene sequence. ITS sequence lengths, as measured from the end of the 16S rRNA gene to the beginning of the 23S rRNA gene, ranged approximately 1 order of magnitude, from 114 base pairs for clone SPOTSAPR01\_5m146 (clone nomenclature as follows: San Pedro Ocean Time Series, month, year, depth, clone, number), which was loosely associated with the 'SAR86' clade within the gamma-Proteobacteria, to 913 base pairs for clone SPOTSAPR01\_5m108, closely associated with the alpha-Proteobacteria *Roseobacter* NAC11-7. Due to this high level of divergence within ITS sequences, only organisms closely related at the 16S rRNA gene level can be compared. However, by comparison to our database, it was possible to place most single ITS sequences retrieved within a 16S rRNA gene phylogenetic framework. When phylogenetic placement of ITS sequences was not possible, the 16S rRNA gene was always sequenced.

In all cases, pairwise comparisons between ITS sequences were significantly ( $p < 0.05$ ) lower than those of the associated 16S rRNA gene sequences. However we did not observe a consistent quantitative relationship, as different taxa displayed ITS sequence divergence at different rates (Fig. 1). At a 16S rRNA gene pairwise similarity above 0.99, the marine cyanobacteria (Fig. 1b) and groups within the alpha- and gamma-Proteobacteria (Fig. 1b,c,e) displayed ITS sequence pairwise similarities below 0.93. This corresponds to the divergence between the low-B/A clades I and II within the *Prochlorococcus* clade as shown in Fig. 2. The majority of pairwise comparisons for ITS sequences were below 0.90 (with many <0.70) when those for corresponding 16S rRNA gene sequences were at or below 0.97. Whilst the SAR11 group does show some ITS sequence pairwise divergence below 0.93 (and even <0.9) within >99% 16S rRNA gene sequence similarity, this clade as a whole displays relatively conserved ITS sequence divergence rates (Fig. 1e,f). Thus, ITS sequence analysis served to place clones in a phylogenetic framework whilst amplifying the microdiversity observed in 16S rRNA gene sequence analysis.

***Prochlorococcus* clade, *Roseobacter* clade and SAR11 clade analysis.** The 3 most abundant ITS sequence types obtained in this study belonged to the *Prochlorococcus* clade, the *Roseobacter* clade, and the SAR11 clade. Within the marine cyanobacteria, sequences associated with the *Prochlorococcus* clade were markedly more prevalent than those associated with the *Synechococcus* clade. Of the 39 ITS sequences related to the *Prochlorococcus* clade, 37 were associated with the *Prochlorococcus* low-B/A clade 1, with all but 2 of these clones originating from the August 2001 and October 2000 5 m libraries (Fig. 2). One sequence each from December and July were associated with the *Prochlorococcus* high-B/A clade 1. No sequences corresponding to the *Prochlorococcus* low-B/A clade II were found. Four sequences from APR01\_5m and 3 sequences from OCT00\_5m were affiliated with the *Synechococcus* clade. Clones SPOTSAPR01\_5m163 and SPOTSAPR01\_5m41 clustered with the *Synechococcus* clade I, whilst the remainder clustered with the *Synechococcus* clade IV. The within-cluster similarity for the marine cyanobacteria subgroups is generally around 98%, whilst the between-cluster mean similarity values are at or below 93%, the similarity between the low-B/A clades I and II. A Chao1 estimate of richness for the *Prochlorococcus* low-B/A clade 1 gives an estimated richness of 132, with coverage of 0.19. Comparisons between ITS sequence divergence and 16S rRNA gene sequence divergence for the marine cyanobacteria are shown in Fig. 1a.

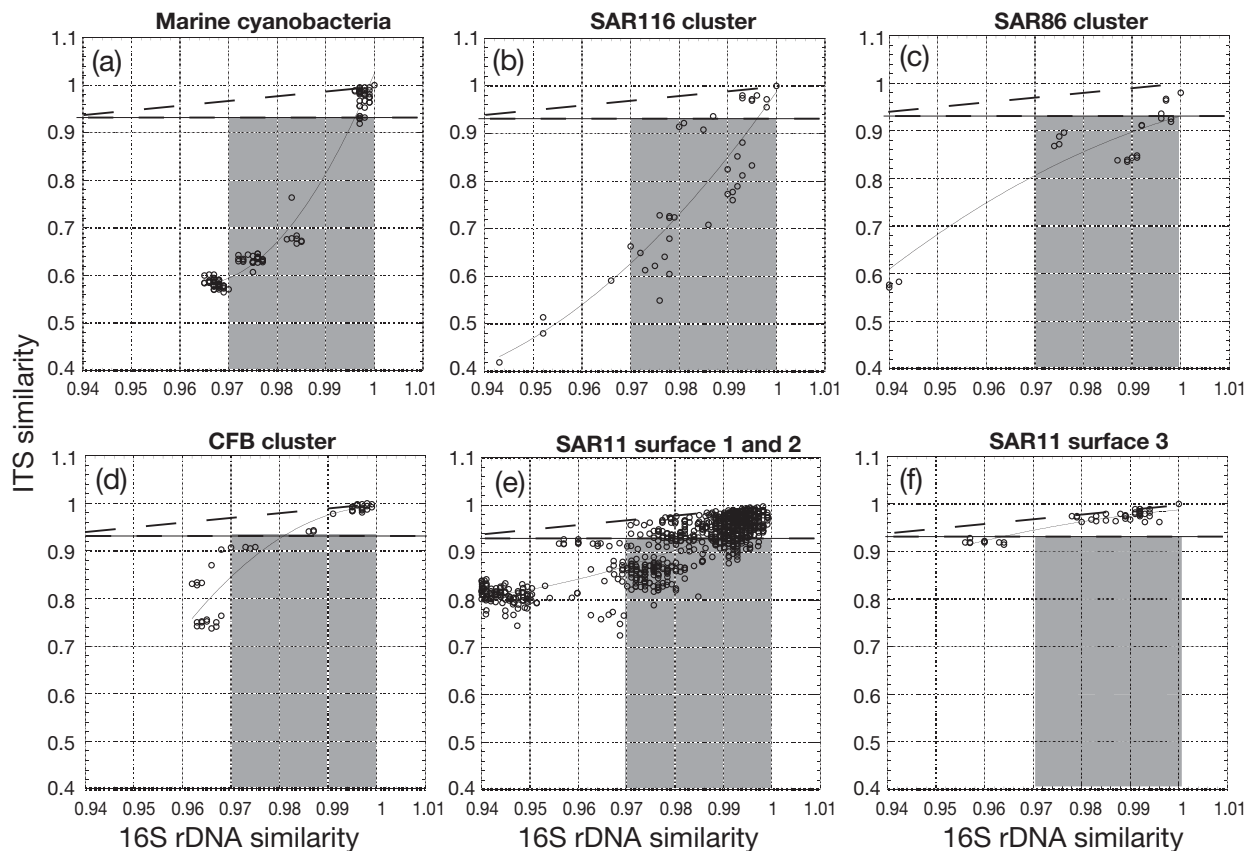


Fig. 1. Comparison of pairwise similarities obtained for full length internal transcribed spacer (ITS) sequences versus full length 16S rRNA gene sequences showing different rates of ITS divergence associated with different taxa. Clones associated with (a) the marine cyanobacteria, (b) the SAR116 cluster, (c) the SAR86 cluster, (d) the *Cytophaga-Flavobacterium-Bacteroides* (CFB) group, (e) the SAR11 surface groups 1, 1(a) and 2, and (f) the SAR11 surface group 3 were used in the analysis. The 1:1 line indicates the expected trajectory if these sequence regions were evolving in tandem. The 0.93 line represents the level below which *Prochlorococcus* isolates have been shown to constitute different ecotypes (Rocap et al. 2002). Grey shading indicates the 'diversity space' where ITS sequences differentiate beyond 93% similarity which is veiled by using 97% 16S rDNA cut off values

The 16S rRNA gene sequences associated with the alpha-Proteobacteria *Roseobacter* clade were obtained only in the April 2001 libraries. These sequences formed a tight cluster, exhibiting >99% 16S rRNA gene sequence identity with the uncultured clone designated *Roseobacter* NAC11-7. Analysis of ITS sequences associated with this cluster identified 2 distinct branches (Fig. 3), with the major branch containing 19 of the 24 sequences. Within these branches, microdiverse relationships are apparent. Within-group average pairwise similarities were calculated at 99.3 and 99.5% for Groups 1 and 2, respectively. The average between-group pairwise sequence similarity was 96.8%. A Chao1 estimate of richness for the *Roseobacter* NAC11-7 clade gives an estimated richness of 87, with coverage of 0.17.

Another alpha-Proteobacteria group, the SAR11 clade, was the most abundant sequence type obtained in this study. Clones associated with this group were prevalent in libraries from all months

examined, with a total of 139 clones retrieved (Fig. 4). Pairwise comparisons between ITS sequence identity and 16S rRNA gene sequence identity for the SAR11 clade are displayed in Fig. 1e,f. The mean similarity between SAR11 clade ITS sequences was 86%, whereas when the alignment was stripped of columns containing gaps (effectively removing addition and deletion events from the analysis) the average pairwise sequence similarity increased to 90%. When only the most conserved region of the spacer (corresponding to the 2 tRNA regions) was considered, the mean similarity was 97%. A Chao1 estimate calculated using ITS sequences from this study gives an estimated total SAR11 clade richness of 793 with coverage of 0.13. As was the case with the marine cyanobacteria, no identical sequences were obtained from 2 different libraries and no SAR11 clade ITS sequences obtained in this study were identical to any of the ~400 SAR11 clade ITS sequences available in the public database.



Fig. 2. Phylogenetic inference obtained from ITS sequences associated with the marine cyanobacterial clade. For clones obtained from the San Pedro Ocean Time Series, nomenclature follows SPOTS(month)(year)\_(depth)(clone number). Group designations follow Rocop et al. (2002)



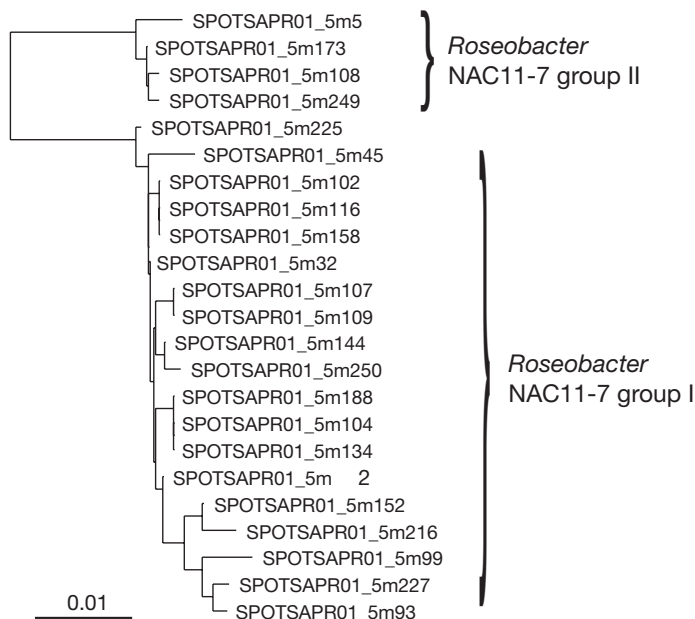


Fig. 3. Phylogenetic inference obtained from ITS sequences associated with the *Roseobacter* NAC11-7 cluster, showing resolution of 2 distinct branches within the >99% similar 16S rRNA gene sequence cluster. For clones obtained from the San Pedro Ocean Time Series, nomenclature follows SPOTS(month)(year)\_(depth)(clone number)

Analysis of SAR11 clade ITS sequences from this study in conjunction with those available from GenBank reveals divergence and clustering of genotypes not only by depth but also by geography (Fig. 4). Four major surface clusters and 3 major deep clusters are evident, with a further cluster containing sequences with mixed affiliations. Differentiation of several of these clusters is supported by bootstrap analysis. By reference to the color codes in Fig. 4 it can be seen that 1 deep cluster, 2 surface clusters and the mixed cluster show no geographic orientation, whilst the remaining 5 clusters are generally dominated by, but importantly not always restricted to, clones from a single location.

## DISCUSSION

### Methodological considerations

Some of the apparent diversity observed using these methods is possibly due to the presence of intragenomic heterogeneity between multiple rRNA operons from single organisms; however, this may not be a significant factor in the marine environment. Specifically concerning the 3 most abundant clone types obtained in this survey (i.e. the SAR11 clade, the *Roseobacter* NAC11-7 clade and the *Prochlorococcus* low-B/A clade 1), 1 cultured representative of the

SAR11 clade, *Pelagibacter ubique* HTCC1062, contains only 1 ribosomal operon (Giovanonni pers. comm. cited in Acinas et al. 2004), whilst *Roseobacter litoralis* DSM6996 has been shown to contain only 1 copy of the 23S rRNA gene (Silke et al. 2004), indicating it too has only 1 rRNA gene operon. Whole genome analysis of *Silicibacter pomeroyii*, an organism related to the *Roseobacter* clade, shows it to have 3 ribosomal operons (Moran et al. 2005). However, these are all identical, even in the highly variable ITS region. Similarly, of the whole genomes available for marine cyanobacteria, the *Prochlorococcus marinus* str. MIT9313 and *Synechococcus* sp. WH8102 have been shown to have multiple rRNA operons, and in both instances the 16S rRNA gene sequences and ITS sequences are identical on the multiple alleles (Palenik et al. 2003, Rocap et al. 2003). Marine isolates may have low copy numbers as a result of their ecological strategies, being relatively slow growers in the oligotrophic environment (Klappenbach et al. 2000). This is consistent with the observation that in our database we did not observe any variation in ITS region length (or sequence) associated with a single 16S rRNA gene sequence.

Therefore, although there are probably marine bacteria with variable rRNA gene operons, there is no evidence to date that these are a significant component in seawater. Effects of multiple operons may be more problematic in more nutrient-rich environments such as soils.

### Microdiversity

Assuming that each unique ITS sequence identified corresponds to a discrete organism, and that temporal mutations in the ITS region are not considerable over the time span of a few years considered herein, the level of microheterogeneity we observed within closely related gene clusters has considerable ecological implications. It suggests that the marine microbial assemblage at the SPOTS station does not consist of organisms derived from a recent common lineage, but rather of a vast array of closely related but discrete co-existing populations. Considering only the sequences obtained in this study, and considering only the 3 most abundant gene clusters—the SAR11 clade (which probably corresponds to multiple species), the *Prochlorococcus* low-B/A clade 1 (which corresponds to 1 'ecotype' of the species) and the NAC11-7 related *Roseobacter* clade (which probably corresponds to 1 species)—the Chao1 diversity indices calculate an estimate of over 1000 unique lineages (at the level of

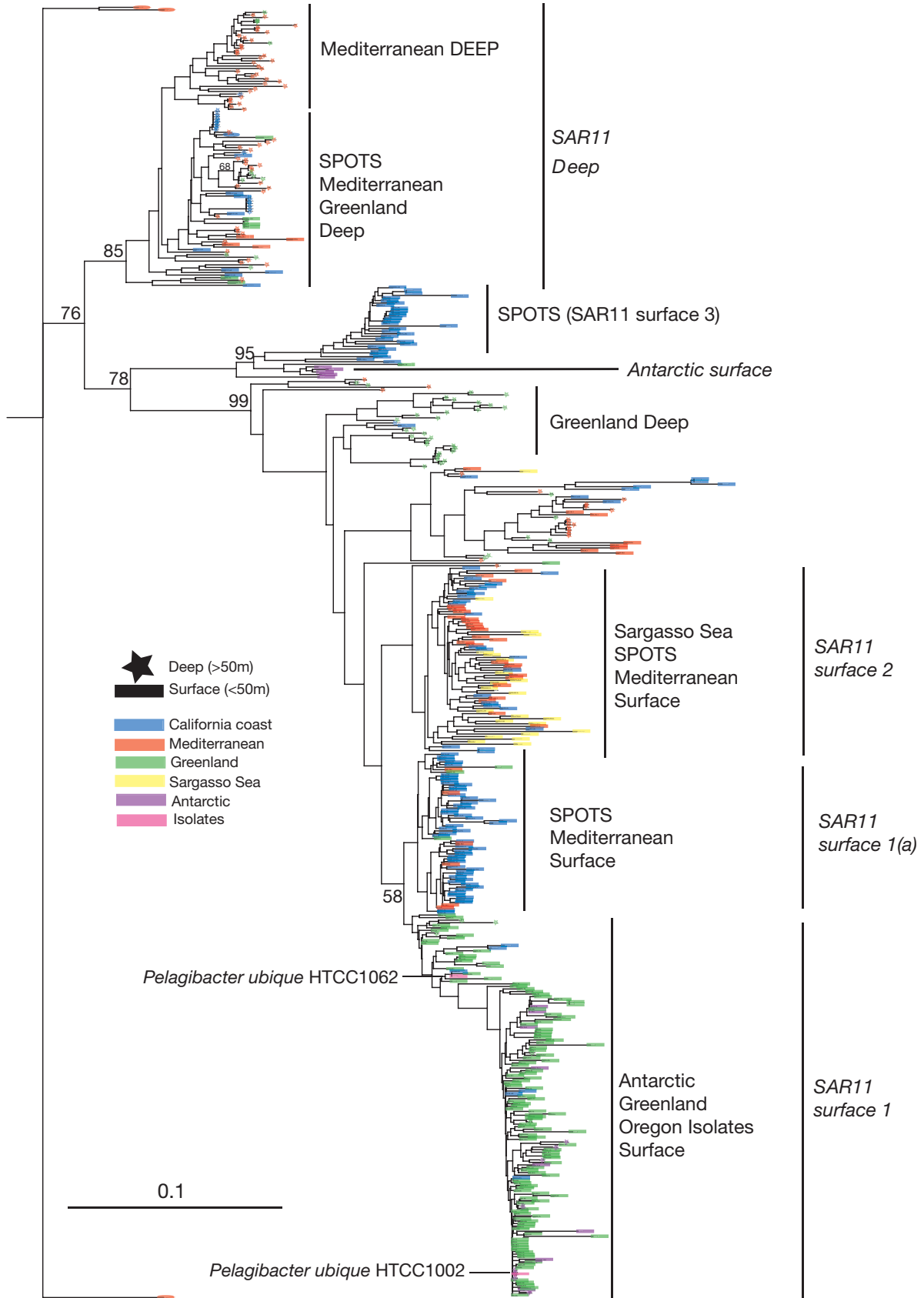


Fig. 4. Phylogenetic inference obtained from ITS sequences associated with the SAR11 clade indicating extensive microdiversity and geographic dominance in several clusters. Color code indicates the geographic origin of the sample site. Significant bootstrap values (>50) for 100 replicates are indicated at the relevant nodes. Group designations in italics are amended from Garcia-Martinez & Rodriguez-Valera (2000)

100% ITS similarity). This finding using the highly variable ITS region provides support for observations made by Acinas et al. (2004) in an estuarine environment and by Venter et al. (2004) in an oligotrophic subtropical environment. However, ITS sequence analysis provides a clearer view of this microdiversity as the few base pair differences observed in 16S rRNA gene sequence comparisons are amplified in the more divergent ITS region. It has been suggested by Acinas et al. (2004) that microdiversity persists in the natural environments because the processes introducing genetic variation are much more prevalent than those which cause a purging of genotypes from the population. Such diversification may be environmentally 'neutral'; however, it may also contribute to a system's biological robustness (which enables a system to maintain its functionalities against external and internal perturbations; Kitano 2004) by confounding processes such as viral attack and protistan grazing (Rocap et al. 2003).

We also repeatedly observed divergence and clustering of ITS sequences even within <1% divergent 16S rRNA gene sequence clusters. Significant genome diversification (and hence potentially phenotypic diversification) has been identified within <1% divergent 16S rRNA gene sequence clusters from a number of environments and using a variety of methods (Keswani & Whitman 2001, Tyson et al. 2004, Thompson et al. 2005), whilst ecologically relevant phenotype differentiation has been observed in organisms with identical or near identical 16S rRNA gene sequences (Ward et al. 1998, Jaspers & Overmann 2004). Furthermore, highly discerning molecular analysis (e.g. multi-locus sequence typing [MLST] or ITS analysis) of cultured isolates has demonstrated that many bacterial 'species' consist of numerous 'ecotypes', often with identical or near identical 16S rRNA gene sequences (e.g. Rocap et al. 2002, Feil et al. 2003). Given this level of phenotype and genotype diversification within closely related organisms, caution should be exercised in extrapolating gene content or metabolic traits of individual isolates to field observations of closely related populations. Highly discerning regions such as the ITS sequence may prove a useful marker for comparison of cultured genome sequences, bacterial artificial chromosome (BAC) libraries and environmental genome constructs to other environments. For example, the *Pelagibacter ubique* genome content may more closely resemble the gene complement of the Antarctic and Greenland SAR11 clade, with which its ITS sequence is most closely associated, rather than the complement of lower latitude associated clones retrieved from the SPOTS station and the Sargasso and Mediterranean Seas. Niche separation corresponding to depth has been identified within the SAR11 clade (Field et al. 1997), and this is supported by both 16S rRNA gene se-

quence and ITS sequence analysis in this and previous studies (Garcia Martinez & Rodriguez-Valera 2000). However, fine-scale phylogeny obtained by ITS sequence analysis highlights the dominance of several surface and deep clusters by sequences originating from a single location, indicating that further niche separation may be occurring. The fact that these clusters also contain a few clones from geographically separate locations suggests that the dominance is attributable to the local environment favoring one 'ecotype' as opposed to differentiation occurring through geographic separation. The samples used in this analysis originated from geographically separated, physically and chemically distinct water bodies. It seems plausible that certain phenotypes would out-compete others in these different environments and lead to dominance in the type of sequence retrieved. Temperature and its environmental correlates may be factors governing the SAR11 clade sequence distributions. Sequences from the very cold (<5°C) Antarctic Ocean and Greenland Sea surface samples cluster together, whilst the highly mixed surface cluster contains only sequences from temperate (SPOTS, Mediterranean) and sub-tropical (Sargasso Sea) locations. Given that the SAR11 strain HTCC1062 contains a rhodopsin-like gene (M. Rappe pers. comm.), latitudinal variations in light regime may also affect this distribution.

We also observed temporal variation in the 3 major groups analyzed. It has been previously shown (Brown et al. 2005) that the SAR11 clade surface 3 cluster displays an increase in relative abundance at the SPOTS station over August to November. Similarly, the appearance of clones associated with the *Roseobacter* NAC11-7 clade solely in the April 2001 libraries, and the abundance of *Prochlorococcus* clade sequences in August and October, are consistent with ARISA data (as well as flow cytometry data for the *Prochlorococcus* clade), showing seasonal variations in these groups over a 4 yr period (Brown et al. 2005). Such variation in temporal abundance is consistent with the hypothesis that these ITS sequence clusters represent functionally differentiated groups that are selected for by temporal variation in environmental parameters.

Using 97% 16S rRNA gene sequence similarity as their definition of an operational taxonomic unit (OTU), Hagström et al. (2002) identified 1117 unique marine-associated bacterial ribotypes represented in the public databases at the time, and concluded that bacterioplankton species richness is relatively low. However, considering the evidence outlined above which points to extensive genomic and phenotypic differentiation within organisms containing >99% similar 16S rRNA gene sequences, it appears 97% 16S rRNA gene sequence similarity is a very conservative OTU definition and that the true bacterioplankton diversity



is much greater than this estimate. In this study, using 100% ITS sequence similarity as our OTU definition, we identified over 1000 unique lineages from only 3 phylogenetic groupings. Whilst 100% ITS sequence analysis almost definitely overestimates the 'functionally differentiated' lineages, ITS sequence clusters may provide a more accurate representation of microbial genotypes than 16S rRNA gene sequence analysis and be a more relevant marker for diversity and biogeography studies (Staley & Gosink 1999).

*Acknowledgements.* We thank I. Hewson, M. Schwabach, J. Steele, X. Hernandez, X. Liang, A. Patel, R. Shimmöler, T. Michaels and the crew of the RV 'Seawatch' for their work on the San Pedro Ocean Time Series. We also acknowledge C. Mahaffey and A. Boal for their considered opinions and advice on the manuscript. This work was supported by NSF Grant MCB0084231 from the Microbial Observatories Program, awarded to J.A.F. and D. Caron. M.V.B. also acknowledges support from the NASA Astrobiology Institute under Cooperative Agreement NNA04CC08A at the Institute for Astronomy (University of Hawaii-Manoa).

#### LITERATURE CITED

- Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 430:551–554
- Brown MV, Schwabach MS, Hewson I, Fuhrman JA (2005) Coupling 16S-ITS rDNA clone libraries and ARISA to show marine microbial diversity; development and application to a time series. *Environ Microbiol* 7:1466–1479
- Feil EJ (2004) Small change: keeping pace with microevolution. *Nature Rev Microbiol* 2:483–495
- Feil EJ, Cooper JE, Grundmann H, Robinson DA and 8 others (2003) How clonal is *Staphylococcus aureus*? *J Bacteriol* 185:3307–3316
- Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, WA
- Field KG, Gordon D, Wright T, Rappe M, Urbach E, Vergin K, Giovanonni SJ (1997) Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* 63:63–70
- 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* 65:4630–4636
- Fuhrman JA, Comeau DE, Hagström Å, Chan AM (1988) Extraction of DNA suitable for molecular biological studies from natural planktonic microorganisms. *Appl Environ Microbiol* 54:1426–1429
- Garcia-Martinez J, Rodriguez-Valera F (2000) Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine archaea of group I. *Mol Ecol* 9:935–948
- Hagström Å, Pommier T, Rohwer F, Simu K, Stolte W, Svensson D, Zweifel UL (2002) Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl Environ Microbiol* 68:3628–3633
- Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* 20:2317–2319
- Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol* 3:reviews0003.1–8
- Hughes JR, Hellermann JJ, Ricketts TH, Bohannan BJM (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* 67:4399–4406
- Jaspers E, Overmann J (2004) Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologicals. *Appl Environ Microbiol* 70:4831–4839
- Keswani J, Whitman WB (2001) Relationship of 16S rRNA sequence similarity to DNA hybridization in prokaryotes. *Int J Syst Evol Microbiol* 51:667–678
- Kitano H (2004) Biological robustness. *Nature Rev Gen* 5:826–837
- Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* 66:1328–1333
- Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, New York, p 115–175
- Ludwig W, Strunk O, Westram R, Richter L and 28 others (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32:1363–1371
- Moran MA, Buchan A, Gonzalez JM, Heidelberg JF and 31 others (2005) Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* 432:910–913
- Palenik B, Brahmamsha B, Larimer FW, Land M and 11 others (2003) The genome of a motile marine *Synechococcus*. *Nature* 424:1037–1042
- Rocap G, Distel DL, Waterbury JB, Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68:1180–1191
- Rocap G, Larimer FW, Lamerdin J, Malfatti S and 20 others (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042–1047
- Schlöter M, Leubhn M, Heulin T, Hartmann A (2000) Ecology and evolution of bacterial microdiversity. *FEMS Microbiol Rev* 24:647–660
- Seurinck S, Verstraete W, Siciliano SD (2003) Use of 16S–23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analysis of *Escherichia coli* isolates to identify nonpoint fecal sources. *Appl Environ Microbiol* 69:4942–4950
- Silke P, Allgaier M, Hocj C, Pauker O, Stackebrandt E, Wagner-Dobler I (2004) Genome organization and localization of the pufLM genes of the photosynthesis reaction center in phylogenetically diverse marine alpha proteobacteria. *Appl Environ Microbiol* 70:3360–3369
- Staley JT, Gosink JJ (1999) Poles apart: biodiversity and biogeography of sea ice bacteria. *Annu Rev Microbiol* 53:189–215
- Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V and 5 others (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* 307:1311–1313
- Tyson GW, Chapman J, Hugenholtz P, Allen EE and 6 others (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:37–43
- Venter JC, Remington K, Heidelberg JF, Halpern AL and 19 others (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74
- Ward D, Ferris MJ, Nold SC, Bateson MM (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol Molec Biol Rev* 62:1353–1370