Environmental Microbiology (2009) **11**(9), 2291–2300 doi:10.1111/j.1462-2920.2009.01954.x

Dynamics of the SAR11 bacterioplankton lineage in relation to environmental conditions in the oligotrophic North Pacific subtropical gyre

Alexander Eiler,[1] Darin H. Hayakawa,[1] Matthew J. Church,[2] David M. Karl[2] and Michael S. Rappé[1]*

[1]*Hawaii Institute of Marine Biology, SOEST, University of Hawaii, Kaneohe, HI 96744, USA.*
[2]*Department of Oceanography, SOEST, University of Hawaii, Honolulu, HI 96822, USA.*

Summary

A quantitative PCR assay for the SAR11 clade of marine *Alphaproteobacteria* was applied to nucleic acids extracted from monthly depth profiles sampled over a 3-year period (2004–2007) at the open-ocean Station ALOHA (A Long-term Oligotrophic Habitat Assessment; 22°45′N, 158°00′W) in the oligotrophic North Pacific Ocean. This analysis revealed a high contribution (averaging 36% of 16S rRNA gene copies) of SAR11 to the total detected 16S rRNA gene copies over depths ranging from the surface layer to 4000 m, and revealed consistent spatial and temporal variation in the relative abundance of SAR11 16S rRNA gene copies. On average, a higher proportion of SAR11 rRNA gene copies were detected in the photic zone (< 175 m depth; mean = 38%) compared with aphotic (> 175 m depth; mean = 30%), and in the winter months compared with the summer (mean = 44% versus 33%, integrated over 175 m depth). Partial least square to latent structure projections identified environmental variables that correlate with variation in the absolute abundance of SAR11, and provided tools for developing a predictive model to explain time and depth-dependent variations in SAR11. Moreover, this information was used to hindcast temporal dynamics of the SAR11 clade between 1997 and 2006 using the existing HOT data set, which suggested that interannual variations in upper ocean SAR11 abundances were related to ocean-climate variability such as the El Niño Southern Oscillation.

Received 30 December, 2008; accepted 11 April, 2009. *For correspondence. E-mail rappe@hawaii.edu; Tel. (+1) 808 236 7464; Fax (+1) 808 236 7443.

Introduction

Recent genetic approaches have underscored the vast biodiversity of microorganisms in the ocean (Venter *et al*., 2004; Rusch *et al*., 2007). Moreover, our knowledge of microbial biogeography continues to increase through the use of DNA-based molecular tools (Giovannoni *et al*., 1990; DeLong, 1992; Fuhrman *et al*., 1992; Rappé and Giovannoni, 2003; Pommier *et al*., 2005; Johnson *et al*., 2006). The widespread use of 16S rRNA gene sequence analyses has identified numerous marine bacterioplankton taxa, of which a genetically diverse clade within the *Alphaproteobacteria* known as SAR11 has been inferred to be among the most abundant (Giovannoni *et al*., 1990; Rappé and Giovannoni, 2003). Bacteria belonging to the SAR11 clade frequently make up 25% or more of the cloned 16S rRNA gene sequences retrieved from seawater samples around the world (Giovannoni and Stingl, 2007). Direct cell counts of SAR11 based on fluorescence *in situ* hybridization have corroborated clone library data and shown that members of the SAR11 clade frequently contribute more than 25% of the total picoplankton cells found in both coastal and open-ocean systems (Morris *et al*., 2002; Malmstrom *et al*., 2004; Alonso and Pernthaler, 2006; Alonso-Saez *et al*., 2007; Vila-Costa *et al*., 2007; Campbell *et al*., 2008; Carlson *et al*., 2009).

The high abundance and global distribution of SAR11 suggest that members of this clade play an important role in organic carbon and nutrient cycling in the ocean. Moreover, these microorganisms appear to contribute substantially to secondary production (as measured by rates of leucine incorporation), and participate in the assimilation of dissolved organic matter substrates, such as amino acids, protein, dimethylsulfoniopropionate and glucose (Malmstrom *et al*., 2004; 2005; Alonso and Pernthaler, 2006; Vila-Costa *et al*., 2007). Recent successes in isolating members of the SAR11 clade, together with comparative analyses of their genomes, has provided insight into their physiology and allowed initial predictions into their ecological roles (Rappé *et al*., 2002; Giovannoni *et al*., 2005a,b; Howard *et al*., 2006; Tripp *et al*., 2008). However, despite increased descriptions of their oceanic distributions, abundances

© 2009 Society for Applied Microbiology and Blackwell Publishing Ltd

and activities, the ecological niches occupied by members of the SAR11 clade, and relationships between these niches and genetic diversity, remain largely unknown.

Oceanographic time series programmes like the Hawaii Ocean Time-series (HOT; Karl and Lukas, 1996) and the Bermuda Atlantic Time-series Study (BATS; Michaels and Knap 1996; Steinberg *et al*., 2001) provide rich data sets on time-dependent biogeochemical and physical variability in subtropical open-ocean ecosystems. Since 1988, HOT has maintained an approximately monthly sampling programme on a suite of oceanographic properties, such as water mass structure, organic and inorganic nutrient concentrations, biological productivity, and particulate matter fluxes at Station ALOHA (Karl and Lukas, 1996). Station ALOHA is situated in one of the largest ocean ecosystems on Earth, the subtropical gyre of the North Pacific Ocean (extending from approximately 15°N to 35°N longitude and 135°E to 135°W latitude and occupying $\sim 2 \times 10^7$ km^2; Sverdrup *et al*., 1946). Oligotrophic gyres represent approximately 75% of the total world oceans. Due to their vast surface area, these gyres play a significant role in global productivity and biogeochemical cycles (Martin *et al*., 1987).

One of the first studies to characterize bacterial 16S rRNA gene sequences from seawater was conducted at Station ALOHA (Schmidt *et al*., 1991), and recent sampling efforts continue to provide insights into microbial diversity and community structure at this location (Konstantinidis and DeLong, 2008; Pham *et al*., 2008; and others). For example, a metagenomic depth profile revealed stratified microbial communities (DeLong *et al*., 2006), and functional gene surveys have revealed temporal and spatial variation in key functional groups of ocean microbes (Church *et al*., 2005; Mincer *et al*., 2007). However, long-term studies of bacterial phylogenetic clades in relation to climate and other environmental variables in the open ocean are limited despite their importance for biogeochemical processes in the largest ecosystem of our planet.

In this study, we utilized a quantitative PCR (QPCR) approach to examine SAR11 dynamics at Station ALOHA. A total of 259 samples, collected over a 3-year period, were analysed to identify depth and time dependent variations in SAR11 16S rRNA gene copy abundances. These data, along with numerous environmental properties, were analysed using multivariate statistics to study environmental forcing on SAR11 bacteria at Station ALOHA. Results from these analyses were used to develop an empirical model of SAR11 abundances in the upper ocean (0–175 m) at this location, which was used to hindcast absolute SAR11 abundances at Station ALOHA from 1997 to 2006.

Results

The HOT oceanographic data set includes measurements of physical, chemical and biotic parameters obtained approximately monthly beginning in 1988. DNA samples used in this study were collected on the monthly cruises to Station ALOHA between 2004 and 2007 (HOT cruises 165–189 and 193). During this period (cruises 165–188, years 2004–2006), depth-integrated (top 100 m) chlorophyll-*a* concentrations varied from 8.7 to 21.9 mg m^{-2}, with peak concentrations occurring in September 2005 (HOT cruise 173), when rates of primary production and inventories of particulate silica were also elevated (Fig. 1). Abundances of non-pigmented (presumably heterotrophic) picoplankton were greatest in the upper 125 m, ranging from 2.0×10^8 to 9.2×10^8 cells l^{-1}, and declining with increasing depth. The depth-dependent decline in the numbers of non-pigmented picoplankton appeared correlated with changes in the concentrations of dissolved organic carbon (DOC) ($R = 0.75$; $P < 0.001$), total phosphate ($R = -0.60$; $P < 0.001$), nitrite plus nitrate ($R = -0.59$; $P < 0.001$) and rates of primary production ($R = 0.51$; $P < 0.001$), based on data from 24 HOT cruises between November 2004 and December 2006. In the upper mixed layer, concentrations of non-pigmented picoplankton correlated with concentrations of particulate phosphorus (PP) ($R = 0.41$; $P < 0.001$), particulate carbon (PC) ($R = 0.39$; $P = 0.001$), the ratio of PC to particulate nitrogen (PN) ($R = 0.52$; $P < 0.001$) and the ratio of PN to PP ($R = -0.35$; $P = 0.003$). However, these apparent correlations were largely the consequence of depth-dependent behaviour in these properties, as reflected by strong autocorrelations of most environmental variables with depth. Depth-integrated (0–100 m) abundances of non-pigmented picoplankton demonstrated only minor temporal fluctuations during this study, ranging between 31.4 and 66.7×10^{12} cells m^{-2}, and did not reveal any significant correlation with other environmental properties measured.

SAR11 16S rRNA gene copy abundances were evaluated using a group-specific quantitative PCR assay from nucleic acid samples collected from 2004 to 2007 at Station ALOHA. These analyses revealed that members of this bacterioplankton clade were abundant throughout the water column (Fig. 2). The relative contribution of SAR11 to total non-pigmented bacterioplankton averaged 38% in the upper 175 m (corresponding approximately to the euphotic zone), decreasing slightly to 22% between 1000 and 4000 m. Correlation analysis only revealed significant covariation between SAR11 and depth when samples from below 700 m were used ($R = -0.404$, $P = 0.003$; Fig. 2). In addition, the relative contribution of SAR11 to the total bacterial 16S rRNA gene copies in the upper 100 m was significantly greater during the winter (December to March;

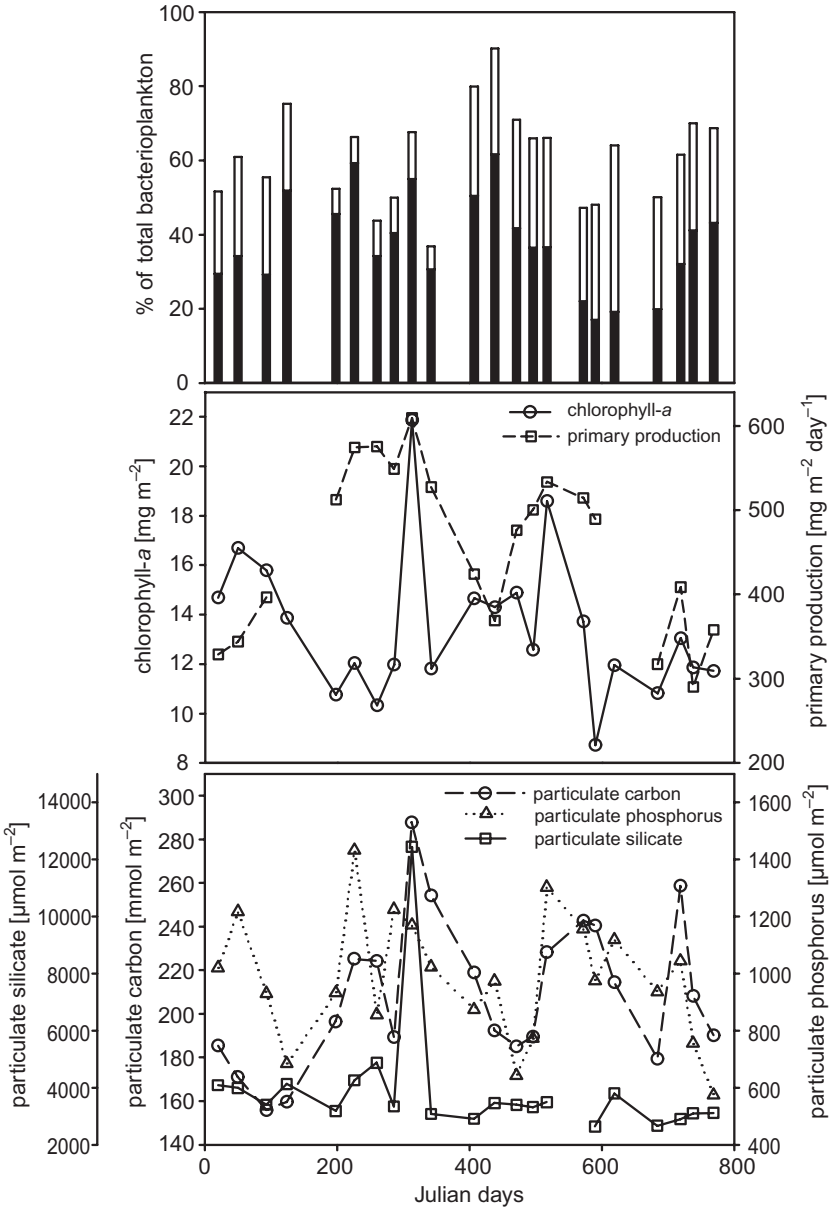


Fig. 1. Depth-integrated (upper 100 m) measurements of (A) the relative contribution of SAR11 to total bacterial 16S rRNA gene copies at Station ALOHA (filled bars), as well as the relative abundance of *Prochlorococcus* measured by flow cytometry (open bars), (B) primary production and chlorophyll-*a*, and (C) particulate carbon, phosphorus and nitrogen. Environmental data are from HOT cruises 165 through 188 (November 2004 through December 2006).

mean = 44%, range = 33–53%), than in the summer (mean = 33%, range = 23–37%) (Kruskal–Wallis test, Hc = 349.5, $P < 0.001$).

Depth-integrated percentages of relative SAR11 16S rRNA gene copy numbers for the upper 100 m ranged between 17% and 60% (Fig. 1), and the combined abundances of SAR11 (estimated by QPCR) and *Prochlorococcus* (enumerated by flow cytometry) constituted more than 50% of the total picoplankton in the upper 100 m during most sampling occasions (Fig. 1). There were no significant relationships between the depth-integrated (0–100 m) relative SAR11 gene abundances and the various biogeochemical and physical properties measured at Station ALOHA, including DOC; primary production; chlorophyll-*a* and other pigments; inorganic nutrients; particulate C, N, P and Si; pH; and temperature.

For cruises where flow cytometric-based enumeration of pigmented and non-pigmented picoplankton cell abundances were available (HOT cruises 165–188, upper 175 m of the water column), absolute SAR11 abundances were calculated using the flow cytometric determinations of total picoplankton cell abundances (excluding pigmented eukaryotic cells), multiplied by the QPCR-derived proportions of SAR11 16S rRNA genes to the total bacterial 16S rRNA gene abundances. Consistent with 16S rRNA gene copy number information available from whole-genome sequences of isolates of the SAR11 clade, a gene copy number of one per cell was used for this

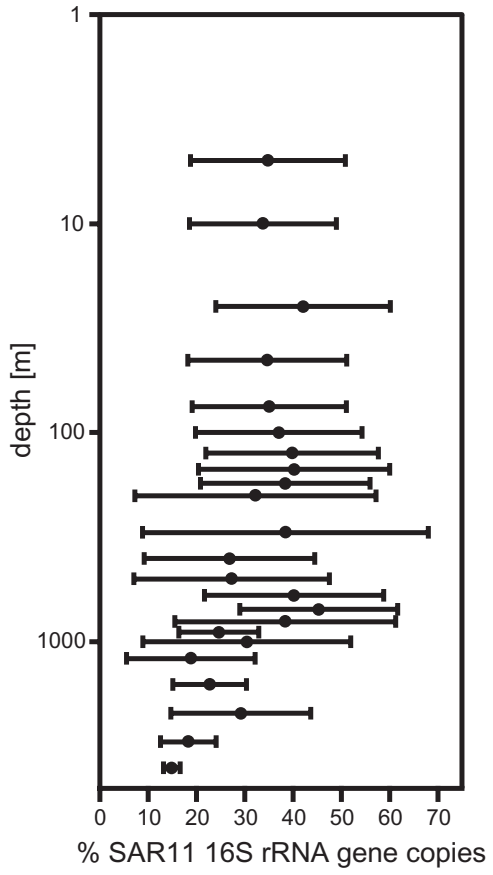


Fig. 2. Average relative contribution of SAR11 to total bacterial 16S rRNA gene copies at Station ALOHA from 2004 to 2007 (HOT cruises 165–188, 189 and 193). Error bars indicate standard deviations.

calculation (Giovannoni *et al*., 2005b; M.S. Rappé and M.J. Huggett, unpubl. results). The resulting SAR11 cell abundance estimates ranged from 0.3×10^5 to 6.3×10^5 cells ml^{-1}. By comparison, total picoplankton cell abundances ranged from 1.2×10^5 to 10×10^5 cells ml^{-1}. Significant correlations were observed between the absolute abundance of SAR11 and the abundances of total non-pigmented cells, *Synechococcus* and *Prochlorococcus* ($P < 0.001$). In addition, the absolute abundance of SAR11 appeared correlated with most environmental data, including significantly positive correlations with DOC ($R = 0.56$; $P < 0.001$), rates of primary production ($R = 0.34$; $P < 0.001$), and particulate C, N, P and Si ($P < 0.01$), and negative correlations with total phosphate ($R = -0.40$; $P < 0.001$) and nitrite plus nitrate ($R = -0.49$; $P < 0.001$). Depth-integrated (0–175 m) abundances of non-pigmented picoplankton and absolute SAR11 abundances were significantly correlated ($R = 0.45$, $P = 0.037$). The integrated (0–175 m) SAR11 stocks correlated with inventories of chlorophyll-*a* ($R = 0.54$, $P = 0.018$), but there were no significant statistical relationships found among SAR11 abundances and other biogeochemical properties measured.

Both the relative and absolute abundances of SAR11 were used to develop a predictive model that described temporal variability in SAR11 based on partial least squares to latent structure (PLS) analyses. However, only the model for absolute abundance provided a satisfactory goodness of fit ($R^2 = 0.56$) and goodness of prediction ($Q^2 = 0.50$). The variables used in the predictive modelling approach included concentrations of DOC, PC, PN, PP, particulate silica (PSi), and chlorophyll-*a*, rates of primary production and the abundance of picocyanobacteria (*Prochlorococcus* and *Synechococcus*). The importance of each variable to predict absolute SAR11 abundances is depicted in Fig. 3A and their scaled and centred coeffi-

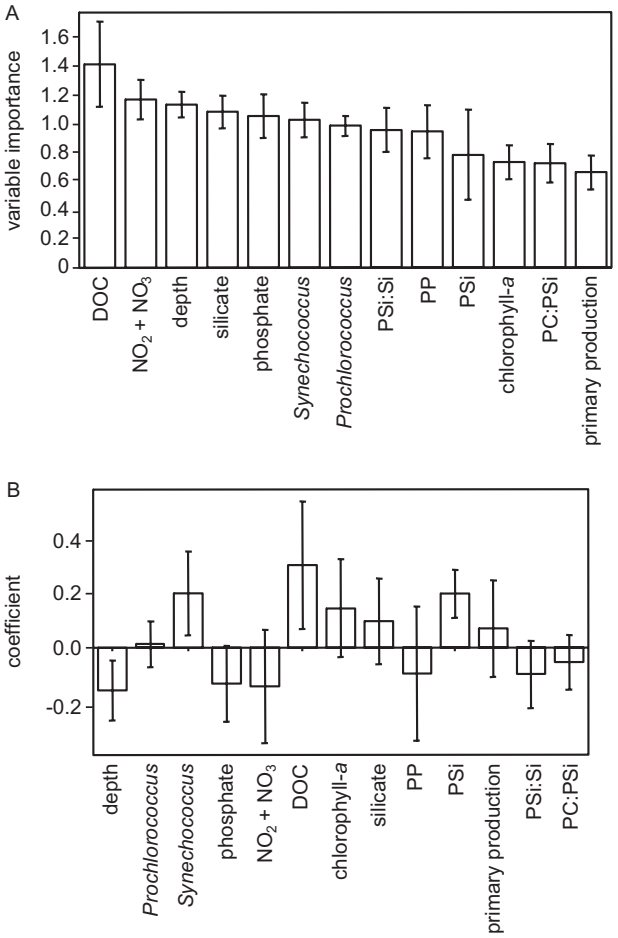


Fig. 3. A. Variable importance plot (VIP) indicating the importance of environmental properties towards the predicting success of the partial least squares (PLS) model for constraining absolute SAR11 abundances.
B. Scaled and centred coefficient plot of environmental variables used in the PLS model predicting absolute SAR11 abundance. Positive values indicate variables that are positively related to the response variable (SAR11 abundance), while negative values indicate variables that diminish SAR11 abundances. Error bars denote 95% confidence intervals derived from jackknifing.

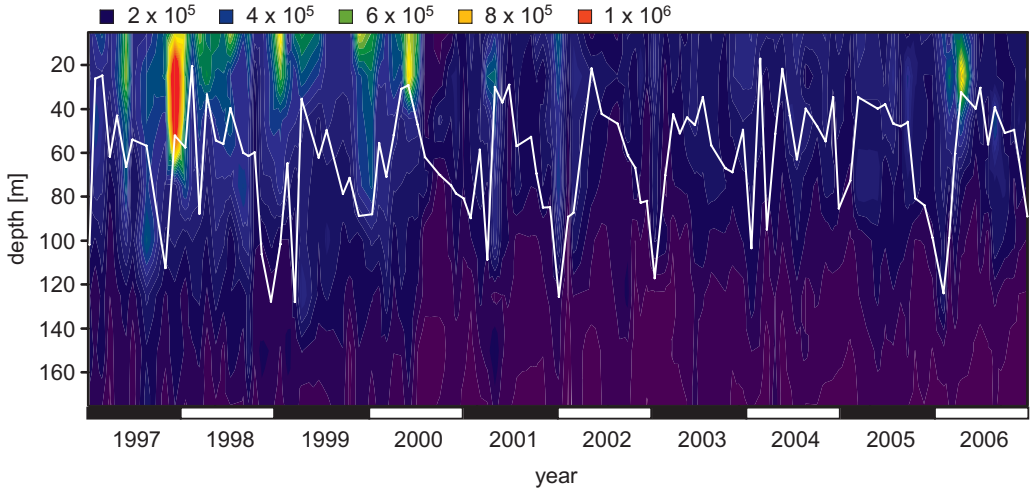


Fig. 4. Predicted absolute SAR11 abundance over 10 years for the upper 175 m at Station ALOHA, calculated with the PLS model (goodness of fit $R^2 = 0.56$; goodness of prediction $Q^2 = 0.50$). The white line indicates the depth of the mixed-layer (salinity pycnocline).

cients are shown in Fig. 3B. The resulting modelled SAR11 abundances were well correlated with measured SAR11 abundances, and the model appeared to capture the prominent patterns of temporal and spatial variability observed in the measured SAR11 gene copy abundances (Fig. S1). The resulting PLS model was subsequently used to hindcast SAR11 distributions over the past 10 years (1997–2006), using measured HOT biogeochemical data acquired during this period (Fig. 4). This hindcast model suggested that SAR11 abundances increased between 1997 and 2000, which are periods that coincided with particularly high variability in El Niño Southern Ocean Oscillation (ENSO) indices (Fig. 4). Regression analysis using yearly averages of the multivariate ENSO index and depth-integrated (0–175 m) SAR11 abundances revealed a positive relationship ($R^2 = 0.70$; $P < 0.001$) (Fig. 5).

Discussion

This study describes the dynamics of the SAR11 clade of bacterioplankton in relation to environmental conditions at Station ALOHA in the North Pacific subtropical gyre. While no clear relationships were detected between the relative abundance of SAR11 16S rRNA gene copies and the numerous habitat properties measured at Station ALOHA, conversion of the relative SAR11 gene abundances to absolute cell numbers revealed several potentially important habitat drivers of absolute SAR11 abundances. For example, correlation analyses revealed significant relationships between absolute SAR11 abundances and PSi concentrations, a proxy for diatom biomass. In September 2005, large increases in upper ocean chlorophyll-*a*, primary production and PSi coincided with increased SAR11 gene abundances. These results suggest that

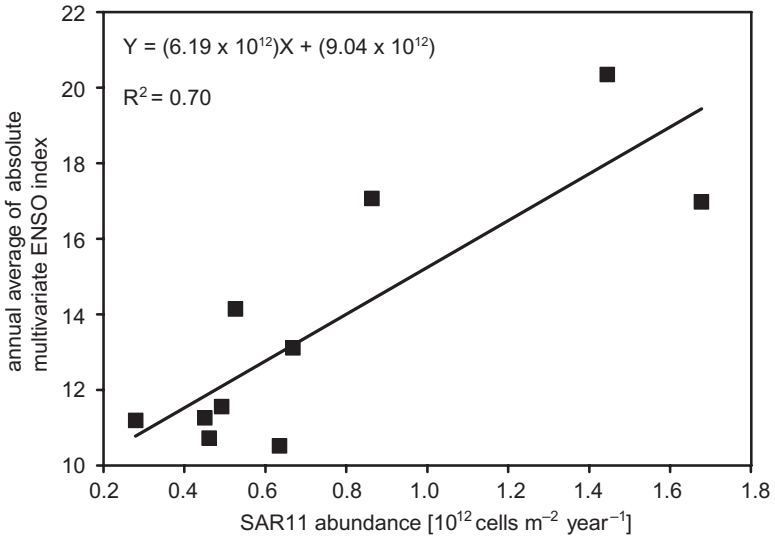


Fig. 5. Regression plot showing the relationship between annual averages of absolute monthly multivariate ENSO (El Niño Southern Oscillation) index and annual averages of depth-integrated (upper 175 m) absolute SAR11 abundance over 10 years (1997–2006). Regression statistics are shown inside the plot.

increases in phytoplankton (specifically, enhancement of diatom biomass) may stimulate SAR11 growth in the oligotrophic ocean. Diatom blooms have been shown to enhance organic matter fluxes, and to play an important role in structuring bacterial community composition and growth in the oligotrophic oceans (Carlson *et al*., 2002; West *et al*., 2008).

In addition to correlation analyses, we used a multivariate statistical procedure to identify environmental properties that may contribute to the observed temporal and spatial variability in SAR11 at Station ALOHA. Partial least squares latent structure analyses resulted in a predictive model of vertical distributions and temporal dynamics of absolute SAR11 abundances. Among the variables that were identified as important predictors of absolute SAR11 dynamics were various indicators of phytoplankton biomass and production, including primary production, DOC, picocyanobacterial abundances and chlorophyll-*a* (Fig. 3A). Variables shown in the variable importance plot (Fig. 3A) can be interpreted as potentially important regulators of absolute SAR11 abundances in this ecosystem; these data emphasize likely temporal and spatial coupling between phytoplankton production and SAR11 growth. Diatom blooms are a regularly occurring feature of the summer months at Station ALOHA (Dore *et al*., 2008), and our results suggest such blooms may increase absolute SAR11 abundance. Interestingly, recent laboratory-based studies with isolated strains of the SAR11 clade have identified a requirement for reduced sulfur, providing additional insight into the dependence of SAR11 on other members of the plankton (Tripp *et al*., 2008). We anticipate that further laboratory-based experimentation will elucidate other linkages between SAR11 activity and phytoplankton production.

Inorganic nutrients such as phosphate and nitrite plus nitrate were negatively correlated with absolute SAR11 abundances, indicating that SAR11 appears to be well adapted to the low nutrient concentrations of surface waters of the North Pacific subtropical gyre. Over the last decade, depth-integrated (0–125 m) phosphate and nitrite plus nitrate stocks averaged 5738 and 5686 μmol m^{-2}, equivalent to 46 and 45 nM respectively. Such low nutrient concentrations are the result of essentially permanent thermal stratification of the upper ocean at Station ALOHA, the strength of which depends on seasonal changes in surface ocean heating-cooling, and may depend on interannual variations in climate forcing (Karl, 2002).

Episodic physical features such as mesoscale eddies have been implicated as nutrient pumps in the oligotrophic ocean (Benitez-Nelson *et al*., 2007). These and other stochastic events, such as storms, seem to drive biotic and abiotic dynamics on short (weeks to month) time scales, whereas climate events like ENSO appear to control interannual and decadal dynamics of the North Pacific subtropical gyre (Karl, 2002; Corno *et al*., 2007). On both inter- and intra-annual scales, observed and modelled abundances of SAR11 showed high variation (Figs 1 and 4, Fig. S1), consistent with observations of spatial and temporal variability in rates of primary production and chlorophyll in this region (Karl, 2002; Letelier *et al*., 2004; Dore *et al*., 2008).

Other studies have also revealed substantial temporal and spatial dynamics of the SAR11 clade (e.g. Morris *et al*., 2002; 2005; Malmstrom *et al*., 2005; Campbell *et al*., 2008; Carlson *et al*., 2009). In the north-western Sargasso Sea, reproducible SAR11 dynamics have been shown to be tightly coupled with seasonal mixing and summer stratification (Morris *et al*., 2005; Carlson *et al*., 2009). There, annual spring phytoplankton blooms result from nutrient entrainment into the photic zone during wintertime convective mixing, with seasonally oligotrophic conditions subsequently intensifying throughout the summer due to stratification of the water column (Steinberg *et al*., 2001). In contrast, mixing in the North Pacific subtropical gyre is much less seasonally dynamic, and as a consequence this ecosystem retains a permanently stratified, oligotrophic water column throughout the year. However, peak primary production at Station ALOHA occurs during the summer months, when nutrients are at their seasonal minimum. Despite these differences, at both locations the absolute abundance of SAR11 increases during summer stratification, peaking in the surface ocean, and decreases during winter and with depth (Carlson *et al*., 2009). Clearly, additional comparisons of SAR11 dynamics at these two locations have the potential to yield further insight into factors controlling the abundance and distribution of this important group of bacteria.

Model-derived absolute abundances of SAR11 suggest that stratification-induced blooms of SAR11 were most pronounced during periods of El Niño-favourable conditions (Fig. 4). Basin-scale climate fluctuations, such as those driven by variations in ENSO and the Pacific Decadal Oscillation, appear to influence upper ocean stratification and vertical nutrient delivery into the euphotic zone at Station ALOHA (Corno *et al*., 2007), and appear to have influenced plankton community composition, rates of primary production and particulate matter export (Karl *et al*., 2001a; Karl, 2002; Corno *et al*., 2007). Our study suggests that the frequency and amplitude of ENSO events are related to, and possibly control, SAR11 abundances in the oligotrophic North Pacific subtropical gyre. The mechanisms of such climate forcing on SAR11 growth are unknown, but results from the PLS analysis hint at climate-dependent variations in phytoplankton and nutrient depletion due to stratification as potential explanations. However, it is important to note that PLS is pri-

marily intended for causal-predictive analysis in situations of high complexity (like biological data) and provides limited potential for hypothesis testing (Wold *et al*., 2001). Additionally, regular PLS only takes into account linear relationships among variables. The model developed in this study also did not fully capture and hindcast the full temporal and spatial variability in biogeochemical properties measured at Station ALOHA (Fig. 4, Fig. S1). Finally, we suspect variation in the structure of SAR11 populations and its differentiation into ecotypes undoubtedly influences the predictive capacity of models like the one developed in this study.

Niche partitioning of SAR11 subtypes has been observed in previous studies (e.g. Field *et al*., 1997; Morris *et al*., 2005; Carlson *et al*., 2009), and it is generally believed that this clade of marine bacteria consists of numerous ecotypes with potentially disparate ecological properties. In addition, recent metagenomic studies have emphasized that SAR11 is a genetically diverse group of organisms (Rusch *et al*., 2007; Wilhelm *et al*., 2007). Several factors have been hypothesized to contribute to the dominance of SAR11 in the euphotic zone, including their possession of the light harvesting pigment proteorhodopsin (Giovannoni *et al*., 2005a) and small genome size (Giovannoni *et al*., 2005b). However, in agreement with previous studies (Morris *et al*., 2002; Carlson *et al*., 2009), we also observed a high contribution of SAR11 bacteria to mesopelagic picoplankton, indicating that the role of light on the physiology of SAR11 cells remains uncertain. Further resolution of populations that correspond to ecologically relevant and biologically uniform entities within the SAR11 clade is a logical next step to increase our knowledge on the biology of SAR11 bacteria and improve predictive models on their population dynamics.

Experimental procedures

Environmental parameters

Samples were obtained from Station ALOHA (22°45′N, 158°W) as part of the ongoing HOT programme (http://hahana.soest.hawaii.edu/index.html;Karl and Lukas 1996). Protocols used for biogeochemical and physical sampling and analyses are described elsewhere (Dore and Karl, 1996; Karl and Lukas, 1996; Letelier *et al*., 1996; Karl *et al*., 2001b; Church *et al*., 2006). Environmental parameters were available for cruises 165 through 188 (November 2004 through December 2006). Monthly multivariate ENSO indexes were obtained from http://www.cdc.noaa.gov/people/klaus.wolter/MEI/table.html and used to calculate annual averages from absolute monthly values.

DNA sampling and extraction

Water samples for nucleic acids were collected from several discrete depths using 12 l polyvinyl chloride bottles attached to a CTD rosette sampler. Samples were collected from 5 ($n = 20$), 10 ($n = 5$), 25 ($n = 21$), 45 ($n = 21$), 75 ($n = 24$), 100 ($n = 28$), 125 ($n = 25$), 150 ($n = 21$), 175 ($n = 27$), 200 ($n = 2$), 300 ($n = 2$), 400 ($n = 6$), 500 ($n = 3$), 600 ($n = 7$), 700 ($n = 2$), 770 ($n = 2$), 800 ($n = 8$), 900 ($n = 2$), 1000 ($n = 8$), 1200 ($n = 2$), 1400 ($n = 1$), 1600 ($n = 2$), 1800 ($n = 1$), 2000 ($n = 6$), 2200 ($n = 1$), 2600 ($n = 1$), 2800 ($n = 1$), 3000 ($n = 4$) and 4000 m ($n = 5$) during HOT cruises 165–189 and 193 (November 2004 through February 2007, and July 2007). Two different procedures were applied on the various cruises. During HOT cruises 165–185, whole seawater samples were subsampled into 4 l, acid-rinsed, polycarbonate bottles of which 1–2 l was filtered through 25 mm diameter, 0.22 μm pore-sized Durapore membrane filters (Millipore, Billerica, MA). Filters were placed in 2.0 ml microcentrifuge tubes containing 500 μl of Tris-EDTA (TE) buffer, immediately quick frozen in liquid nitrogen, and stored at –80°C until extraction. DNA was extracted following the protocols described in Church and colleagues (2005). DNA extracts were purified using the Qiagen DNeasy Kit (Qiagen, Valencia, CA) following the manufacturer's protocols, and stored at –80°C. During cruises 185–189 and 193, 0.5–4 l of seawater samples was filtered onto 13 mm diameter, 0.2 μm pore-sized polyethersulfone filters (Supor-200; PALL, Ann Arbor, MI). The membrane filters were subsequently stored in 2 ml screw-cap tubes containing 250 μl of DNA lysis buffer [20 mM Tris-HCl, 2 mM EDTA, 1.2% Triton X-100, 2% lysozyme (w/v), pH 8] at –80°C until further processing. Total community genomic DNA was extracted from the microbial communities contained on the membrane filters using the Qiagen DNeasy 96 Tissue Kit following the manufacturer's recommended protocol for bacteria.

Quantitative PCR

QPCR reactions were prepared using the SensiMix dT kit (Quantace, London, Great Britain) in 0.1 ml MicroAmp 96-well plates (Applied Biosystems, Foster City, CA) and a StepOnePlus PCR system (Applied Biosystems). Reactions contained 1× SensiMix dT master mix, 250 nM of each primer (final concentration), 0.2–15 ng of template (DNA), 1× SYBR Green I solution and UV-treated 0.1 μm filter sterilized water in 20 μl reactions. For each sample, two individual replicated QPCRs were performed using different primer pairs. Primers used for amplification of SAR11 clade-specific 16S rRNA genes were 433f (5′-CTCTTTCGTC GGGGAAGAAA-3′) and 588r (5′-CCACCTACGWGCTCTT TAAGC-3′) (Suzuki *et al*., 2001). PCR was performed using a thermocycling programme consisting of an initial step of 5 min at 95°C, followed by 45 cycles of 95°C for 15 s (denaturing), 58°C for 45 s (annealing), 72°C for 45 s (extension) and 78°C for 15 s of SYBR Green fluorescence signal acquisition. Total bacterial 16S rRNA gene copy numbers were assessed using primers 27F-B (5′-AGRGTTYGATYMTGGC TCAG-3′) (Vergin *et al*., 1998) and 519R (5′-GWATTACC GCGGCKGCTG-3′) (Lane, 1991), and the same PCR assay as described above, with the exception that annealing and acquisition temperatures of 55°C and 77°C were used respectively.

Results were analysed with StepOne software Version 2.0 (Applied Biosystems). For all samples, melting profiles were

Table 1. Characteristics of the QPCR assays used in this study, including average regression coefficients (r^2), amplification efficiencies (E) and slope of the standard curves.

QPCR assay	r^2	E	Slope
Bacteria	0.994 (0.002)	72.9 (6.8)	−4.24 (0.30)
SAR11	0.998 (0.003)	90.0 (6.2)	−3.60 (0.18)

Standard deviations are shown in parentheses. A total of eight QPCR runs for each primer pair were required to analyse all samples in replicates.

obtained to check if the length of the resulting PCR product was homogenous. Positive controls and titres for the QPCR assay consisted of plasmid DNA containing a 16S rRNA gene and internal transcribed spacer (2.1 kb) from SAR11 strain HIMB104, isolated from Kaneohe Bay, Hawaii (M.S. Rappé and M.L. Brandon, in review). Plasmid DNA was diluted in 10-fold steps from 1.4 ng to 0.14 pg per reaction, and duplicates of each dilution were analysed in parallel with the environmental samples. Negative controls consisted of QPCR reactions lacking template DNA. Simple regressions of the cycle threshold values for the DNA standards versus the log-transformed starting copy number were plotted for both the general bacterial primer pair (total number of 16S rRNA genes) and the 16S rRNA genes targeted by the SAR11-specific primer pair. Since 96-well plates were used, eight runs for each primer pair (including 10 standards and two blanks in each run) were required to process all samples. To provide information about reproducibility and sensitivity of all QPCR runs, mean and standard deviation for QPCR statistics, like r^2 value, amplification efficiency and slope, were reported (Table 1). Absolute numbers of SAR11 were estimated by multiplying the bacterial cell abundance, determined by flow cytometry, with the relative contribution of the SAR11 gene copy number to the total 16S rRNA gene copy number.

Clone library construction and analysis

In order to confirm the specificity of the SAR11 QPCR assay, QPCR products from four different samples (cruise 170–45 m, cruise 174–75 m, cruise 177–25 m and cruise 186–2000 m) were cloned using the pGem Easy cloning kit, following protocols recommended by the manufacturer (Promega, Madison, WI). At least six clones from each clone library were sequenced on an ABI 3700 96-capillary sequencer (Applied Biosystems) using primer M13f and the BigDye terminator kit version 3.1 (Applied Biosystems). High-quality DNA sequences spanning the full length of the cloned amplicons (*c.* 250 bp) were compared with the GenBank sequence database via BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). All 24 of the resulting sequences were 99–100% similar to SAR11-related 16S rRNA genes in the GenBank sequence database.

Statistical analyses and modelling

Three stages of statistical analyses were conducted that ultimately resulted in predictive models for SAR11 abundance at Station ALOHA. First, Spearman correlation analyses were performed with the environmental parameters and relative contribution of SAR11 16S rRNA copy numbers as well as estimated absolute SAR11 abundances (Statistica, StatSoft, Tulsa, OK). These correlation analyses identified several relationships between the measured environmental variables and SAR11 abundances, and highlighted several autocorrelations between the environmental variables. Second, samples were grouped by depth, by photic zone (above 100 m) and below, and by season (winter was defined by December through March, summer by June through September, and spring and autumn represent the months in between). Significant differences between groups were tested using Kruskal–Wallis test with Bonferroni correction (Statistica). Third, models were developed based on partial least square projection to latent structures (PLS) with autoscaling in Simca-Pþ version 11.0 (Umetrics AB, Umeå, Sweden; Wold *et al.*, 2001). The environmental variables were log-transformed, when appropriate, to obtain normality before analysis. Environmental variables demonstrating little influence on the model were iteratively excluded from the analysis based on the variable importance plot until the goodness of prediction (Q^2) reached a maximum. The goodness of prediction is obtained after cross-validation (Wold *et al.*, 2001) and is important to avoid overfit. Cross-validation excludes a subset of the data; the remaining data set is used to fit a model (training data set) that is used to predict the validation data sets. This was done by dividing the data set into seven random subsets and by using each combination of six subsets as the training data set to fit the remaining subset (validation data set), until all observations were excluded once. The goodness of prediction was than calculated using the following equation: $Q^2 = 1 - \Sigma(y_i - y_{i,\,pred})^2/\Sigma(y_i - y_{mean})^2$ where y_i is the observed SAR11 abundance, $y_{i,\,pred}$ is the predicted SAR11 abundance and y_{mean} is the mean of the observed SAR11 abundances.

Jackknifing was used to estimate the confidence interval of the variables coefficient in predicting SAR11 abundance. Leaving out one observation at a time from the data set, the model was fitted and the coefficients of variables used to predict SAR11 abundances were calculated. The obtained coefficient of each variable from all runs was then used to estimate the 95% confidence interval. Unlike other multivariate statistical methods such as multiple linear regression, PLS analyses data with strongly collinear (autocorrelated), noisy and numerous X-variables. X-variables required for the model were downloaded from the online HOT programme data repository (http://hahana.soest.hawaii.edu/hot/hot_jgofs.html) and used to hindcast SAR11 abundances using Simca-Pþ version 11.0.

Acknowledgements

The authors thank the captains and crews of the many research vessels that supported the HOT programme from 1996 to 2007. We thank J. Becker, M. Miller, T. Campbell, S. Yeo, J. Coloma, N. Wagoner, E. Hambleton and all HOT personnel for assistance in sampling. This research was supported by funding from the Center for Microbial Oceanography: Research and Education (NSF Science and Technology Center Award EF-0424599) and NSF award OCE-0326616 to D.M.K. This is SOEST contribution 7679 and HIMB contribution 1344.

References

Alonso, C., and Pernthaler, J. (2006) Roseobacter and SAR11 dominate microbial glucose uptake in coastal North Sea waters. *Environ Microbiol* **8:** 2022–2030.

Alonso-Saez, L., Balague, V., Sa, E.L., Sanchez, O., Gonzalez, J.M., Pinhassi, J., *et al.* (2007) Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol Ecol* **60:** 98–112.

Benitez-Nelson, C.R., Bidigare, R.R., Dickey, T.D., Landry, M.R., Leonard, C.L., Brown, S.L., *et al.* (2007) Mesoscale eddies drive increased silica export in the subtropical Pacific Ocean. *Science* **316:** 1017–1021.

Campbell, B.J., Waidner, L.A., Cottrell. M.T., and Kirchman, D.L. (2008) Abundant proteorhodopsin genes in the North Atlantic Ocean. *Environ Microbiol* **10:** 99–109.

Carlson, C.A., Giovannoni, S.J., Hansell, D.A., Goldberg, S.J., Parsons, R., Otero, M.P., *et al.* (2002) The effect of nutrient amendments on bacterioplankton production, community structure and DOC utilization in the northwestern Sargasso Sea. *Aquat Microb Ecol* **30:** 19–36.

Carlson, C.A., Morris, R., Parsons, R., Treusch, A.H., Giovannoni, S.J., and Vergin, K. (2009) Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* **3:** 283–295.

Church, M.J., Jenkins, B.D., Karl, D.M., and Zehr, J.P. (2005) Vertical distributions of nitrogen-fixing phylotypes at Station ALOHA in the oligotrophic North Pacific Ocean. *Aquat Microb Ecol* **38:** 3–14.

Church, M.J., Ducklow, H.W., Letelier, R.M., and Karl, D.M. (2006) Temporal and vertical dynamics in picoplankton photoheterotrophic production in the subtropical North Pacific Ocean. *Aquat Microb Ecol* **45:** 41–53.

Corno, G., Karl, D.M., Church, M.J., Letelier, R.M., Lukas, R., Bidigare, R.R., and Abbott, M.R. (2007) Impact of climate forcing on ecosystem processes in the North Pacific Subtropical Gyre. *J Geophys Res* **112:** C04021.

DeLong, E.F. (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci USA* **89:** 5685–5689.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496–503.

Dore, J.E., and Karl, D.M. (1996) Nitrite distributions and dynamics at station ALOHA. *Deep Sea Res Part II* **43:** 385–402.

Dore, J.E., Letelier, R.M., Church, M.J., Lukas, R., and Karl, D.M. (2008) Summer phytoplankton blooms in the oligotrophic North Pacific Subtropical Gyre: historical perspective and recent observations. *Prog Oceanogr* **76:** 2–38.

Field, K.G., Gordon, D., Wright, T., Rappé, M., Urbach, E., Vergin, K., and Giovannoni, S.J. (1997) Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* **63:** 63–70.

Fuhrman, J.A., McCallum, K., and Davis, A.A. (1992) Novel major archaebacterial group from marine plankton. *Nature* **356:** 148–149.

Giovannoni, S.J., and Stingl, U. (2007) The importance of culturing bacterioplankton in the 'omics' age. *Nature Rev Microbiol* **5:** 820–826.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G. (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345:** 60–63.

Giovannoni, S.J., Bibbs, L., Cho, J.-C., Stapels, M.D., Desiderio, R., Vergin, K.L., *et al.* (2005a) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438:** 82–85.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., *et al.* (2005b) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309:** 1242–1245.

Howard, E.C., Henriksen, J.R., Buchan, A., Reisch, C.R., Bürgmann, H., Welsh, R., *et al.* (2006) Bacterial taxa that limit sulfur flux from the Ocean. *Science* **314:** 649–652.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S., and Chisholm, S.W. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311:** 1737–1740.

Karl, D.M. (2002) Nutrient dynamics in the deep blue sea. *Trends Microbiol* **10:** 410–418.

Karl, D.M., and Lukas, R. (1996) The Hawaii Ocean Time-series (HOT) program: background, rationale and field implementation. *Deep Sea Res Part II* **43:** 129–156.

Karl, D.M., Bidigare, R.R., and Letelier, R.M. (2001a) Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: the domain shift hypothesis. *Deep Sea Res Part II* **48:** 1449–1470.

Karl, D.M., Bjorkman, K.M., Dore, J.E., Fujieki, L., Hebel, D.V., Houlihan, T., *et al.* (2001b) Ecological nitrogen-to-phosphorus stoichiometry at Station ALOHA. *Deep Sea Res Part II* **48:** 1529–1566.

Konstantinidis, K.T., and DeLong, E.F. (2008) Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *ISME J* **2:** 1052–1065.

Lane, D.J. (1991) 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics.* Stackebrandt, E., and Goodfellow, M. (eds). Chichester, UK: Wiley and Sons, pp. 115–175.

Letelier, R.M., Dore, J.E., Winn, C.D., and Karl, D.M. (1996) Seasonal and interannual variations in photosynthetic carbon assimilation at Station ALOHA. *Deep Sea Res Part II* **43:** 467–487.

Letelier, R.M., Karl, D.M., Abbott, M.R., and Bidigare, R.R. (2004) Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre. *Limnol Oceanogr* **49:** 508–519.

Malmstrom, R.R., Kiene, R.P., Cottrell, M.T., and Kirchman, D.L. (2004) Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic Ocean. *Appl Environ Microbiol* **70:** 4129–4135.

Malmstrom, R.R., Cottrell, M.T., Elifantz, H., and Kirchman, D.L. (2005) Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Appl Environ Microbiol* **71:** 2979–2986.

Martin, J.H., Knauer, G.A., Karl, D.M., and Broenkow, W.W. (1987) VERTEX: carbon cycling in the northeast Pacific. *Deep-Sea Res* **34:** 267–286.

Michaels, A.F., and Knap, A.H. (1996) Overview of the US JGOFS Bermuda Atlantic Time-Series Study and the Hydrostation S program. *Deep Sea Res Part II* **43:** 157–198.

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., and DeLong, E.F. (2007) Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol* **9:** 1162–1175.

Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806–810.

Morris, R.M., Vergin, K.L., Cho, J.-C., Rappé, M.S., Carlson, C.A., and Giovannoni, S.J. (2005) Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-Series Study site. *Limnol Oceanogr* **50:** 1687–1696.

Pham, V.D., Konstantinidis, K.T., Palden, T., and DeLong, E.F. (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environ Microbiol* **10:** 2313–2330.

Pommier, T., Pinhassi, J., and Hagström, Å. (2005) Biogeographic analysis of ribosomal RNA clusters from marine bacterioplankton. *Aquat Microb Ecol* **41:** 79–89.

Rappé, M.S., and Giovannoni, S.J. (2003) The uncultured microbial majority. *Ann Rev Microbiol* **57:** 369–394.

Rappé, M.S., Connon, S.A., Vergin, K.L., and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418:** 630–633.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al.* (2007) The *Sorcerer II* global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5:** e77.

Schmidt, T.M., DeLong, E.F., and Pace, N.R. (1991) Analysis of a marine picoplankton community by 16S ribosomal-RNA cloning and sequencing. *J Bacteriol* **173:** 4371–4378.

Steinberg, D.K., Carlson, C.A., Bates, N.R., Johnson, R.J., Michaels, A.F., and Knap, A.H. (2001) Overview of the US JGOFS Bermuda Atlantic Time-Series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. *Deep Sea Res Part II* **48:** 1405–1447.

Suzuki, M.T., Preston, C.M., Chavez, F.P., and DeLong, E.F. (2001) Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay. *Aquat Microb Ecol* **24:** 117–127.

Sverdrup, H.U., Johnson, M.W., and Fleming, R.H. (1946) *The Oceans*. Englewood Cliffs, NJ, USA: Prentice Hall.

Tripp, H.J., Kitner, J.B., Schwalbach, M.S., Dacey, J.W.H., Wilhelm, L.J., and Giovannoni, S.J. (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452:** 741–744.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Vergin, K.L., Urbach, E., Stein, J.L., DeLong, E.F., Lanoil, B.D., and Giovannoni, S.J. (1998) Screening a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order *Planctomycetales*. *Appl Environ Microbiol* **64:** 3075–3078.

Vila-Costa, M., Pinhassi, J., Alonso, C., Pernthaler, J., and Simo, R. (2007) An annual cycle of dimethylsulfoniopropionate-sulfur and leucine assimilating bacterioplankton in the coastal NW Mediterranean. *Environ Microbiol* **9:** 2451–2463.

West, N.J., Obernosterer, I., Zemb, O., and Lebaron, P. (2008) Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. *Environ Microbiol* **10:** 738–756.

Wilhelm, L.J., Tripp, H.J., Givan, S.A., Smith, D.P., and Giovannoni, S.J. (2007) Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct* **2:** 27.

Wold, S., Sjöström, M., and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Laboratory Syst* **58:** 109–130.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. (A) Plot of observed vs. predicted absolute SAR11 abundances at Station ALOHA, calculated with the PLS model (goodness of fit $R^2 = 0.56$; goodness of prediction $Q^2 = 0.50$). (B) Contour plots of observed and predicted absolute SAR11 abundances at Station ALOHA.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.