Final Report: DOE-Biological Ocean Margins Program.
Microbial Ecology of Denitrifying Bacteria in the Coastal Ocean.

DE-FG02- 00ER62978-2000-2003
Lee J. Kerkhof- **PI**
Institute for Marine and Coastal Sciences
phone: (848) 932-3419 fax: (732) 932-8520 email: lkerkhof@rutgers.edu

OBJECTIVES

The initial goal of the proposed work was to use the nitrous oxide reductase gene (*nosZ*) for characterizing denitrifying bacteria in coastal samples off New Jersey. Unfortunately, the priming sets we designed for *nosZ* were found to mis-prime when the di-deoxynucleotides we had used in the PCR reaction were no longer commercially available and replaced using a significantly cleaner dNTP mix from Roche. Our analytical problems using *nosZ* were only resolved near the end of the project period. As such, we initiated discussions with the DOE program manager and changed the focus of our research to providing a comprehensive study of the bacterioplankton populations off the coast of New Jersey near the Rutgers University marine field station using terminal restriction fragment polymorphism analysis (TRFLP) coupled to 16S rRNA genes for large data set studies.

Our three revised objectives to this study became:

1) to describe bacterioplankton population dynamics in the Mid Atlantic Bight (figure 1) using TRFLP analysis of 16S rRNA genes.
2) to determine whether spatial and temporal factors are driving bacterioplankton community dynamics in the MAB using monthly samping along our transect line over a 2 year period.
3) to identify dominant members of a coastal bacterioplankton population by clonal library analysis of 16S rDNA genes and sequencing of PCR product corresponding to specific TRFLP peaks in the data set.

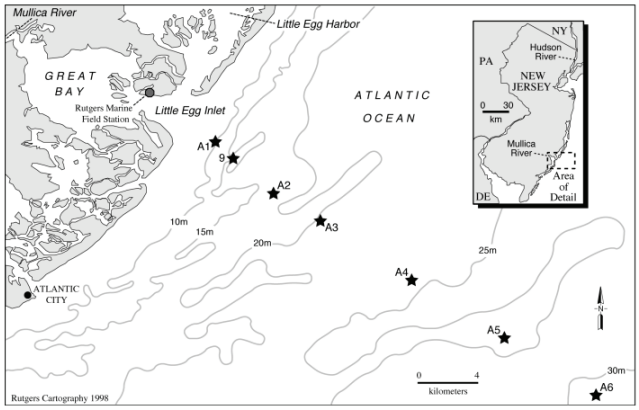

Figure 1. Study site off the coast of New Jersey in the Mid-Atlantic Bight.

Results

Although open ocean time-series sites have been areas of microbial research for years, relatively little was known about the population dynamics of bacterioplankton communities in the coastal ocean on kilometer spatial and seasonal temporal scales. To gain a better understanding of microbial community variability, monthly samples of bacterial biomass were collected in 1995–1996 along a 34-km transect near the Long-Term Ecosystem Observatory (LEO-15) off the New Jersey coast. Surface and bottom sampling was performed at seven stations along a transect line with depths ranging from 1 to 35m (n=178). The data revealed distinct temporal patterns among the bacterioplankton communities in the Mid-Atlantic Bight rather than grouping by sample location or depth (figure 2-next page). Principal components analysis models supported the temporal patterns. In addition, partial least squares regression modeling could not discern a

significant correlation from traditional oceanographic physical and phytoplankton nutrient parameters on overall bacterial community variability patterns at LEO-15. These results suggest factors not traditionally measured during oceanographic studies are structuring coastal microbial communities (Nelson et al., 2008).

In this seasonal study, we initially attempted to determine whether time or space was the primary driver for organizing the microbial community in the Mid-Atlantic Bight. Although substantial gradients in physical and phytoplankton nutrient parameters were observed along our 34km transect off coastal New Jersey, the primary forcing factor for the coastal bacteria appeared to be temporal. The formation of major seasonal and minor spatial groups in this study was demonstrated by UPGMA analysis using Sorensen's similarity indices based on the presence or absence of TRFs. Additionally, the results from multivariate statistical analyses agreed well with UPGMA data with PCA and PLS models confirming the seasonal effect on bacterioplankton communities based on TRFLP profiles (Nelson et al., 2008).

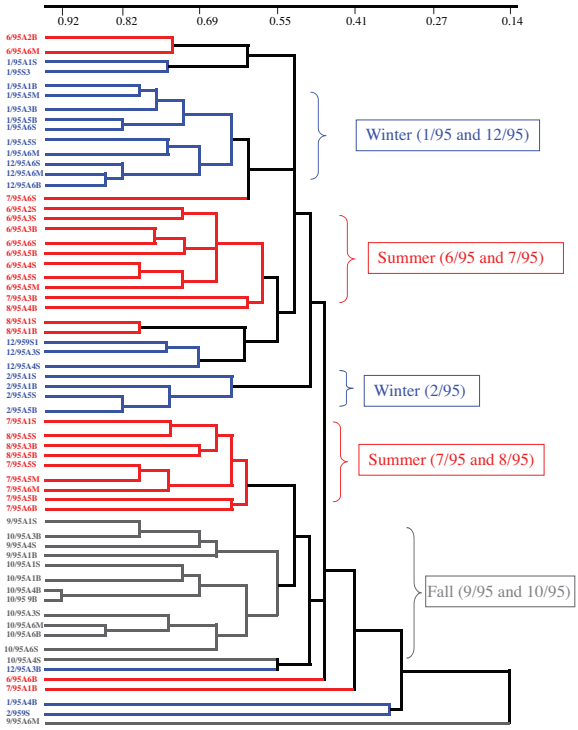

Figure 2. Relationship between electron transport system (ETS) and vertical attenuation coefficient (Kd) for cruise OriPEx VIII.

The results of our MAB study indicate that there is no one single regulator that is responsible for bacterial community dynamics in sediments (Scala & Kerkhof, 2000). Some of these seasonal populations drivers of bacteria at the LEO-15 site could include: changes in overall quality or quantity of organic matter, shifts in grazing pressure for distinct microbial populations, or advection of larger microbial biomes into the study site. At present, there is evidence of little to no overall change in DOC quantity or quality in the Mid-Atlantic Bight with season or over the depths encountered during this study (Aluwihare et al., 2002; Del Vecchio & Blough, 2004). However, none of these hypotheses has been directly tested as a driver for structuring microbial communities in the Mid-Atlantic Bight. Our results underscore the need for a wider array of chemical and physical data collection to understand bacterioplankton community dynamics and illustrate how time series analysis of larger spatial regions can lead to testable hypotheses to discern the mechanisms controlling bacterial community structure and function.

Published Manuscripts funded in whole or in part by the grant include:

Nelson, J. D., Boehme, S. E., Reimers, C. E., Sherrell, R. M., and L. J. Kerkhof. 2008. Temporal patterns of microbial community structure in the Mid-Atlantic Bight. *FEMS Micro. Ecol.* 65: 484-493

Mills, H. J., Hunter, E., Humphrys, M., Kerkhof, L., McGuinness, L., Huettel, M. and J. E. Kostka. 2008. Characterization of Nitrifying, Denitrifying, and Overall Bacterial

Communities in Permeable Marine Sediments of the Northeastern Gulf of Mexico. *Appl. Environ. Microbiol.* 74: 4440–4453

Babcock, D.A., B. Wawrik, J.H. Paul, L. McGuinness, L.J. Kerkhof. 2007. Rapid screening of a large insert BAC library for specific 16S rRNA genes using TRFLP. *J. Micro. Methods* 71: 156–161

Perez-Jimenez, J. R. and L. J. Kerkhof. 2005. Phylogeography of sulfate-reducing bacteria among disturbed sediments disclosed by analysis of the dissimilatory sulfite reductase genes (*dsrAB*). *Appl. Environ. Microbiol.* 71: 1004-1011

Corredor, J. Wawrik, B., Paul, J. Tran, H., Kerkhof, L, Lopez, J., Dieppa, A. and O. Cardenas. 2004. Geochemical Rate-RNA integration study: Ribulose 1,5 Bisphosphate Carboxylase/ Oxygenase Gene Transcription and Photosynthetic Capacity of Planktonic Photoautotrophs. *Appl. Environ. Microbiol.* 70: 5459–5468

Vetriani, C, Tran, H. V., and L. Kerkhof. 2003. Fingerprinting Microbial Assemblages from the Oxic/Anoxic Chemocline of the Black Sea. *Appl. Environ. Microbiol.* 2003. 69: 6481-6488

Published Manuscripts submitted or in prep include:

J. D. Nelson, D. Babcock, and L. Kerkhof. Dominant Bacteria in the Mid-Atlantic Bight. Draft ready for submission to BMC Microbiology

D. Babcock and L. Kerkhof. Sequence Analysis of Large-insert BAC clones from Dominant Microorganisms in the Mid-Atlantic Bight. Draft ready for submission to FEMS Microbiol. Ecol.

L. J. Kerkhof L. M. McGuinness, and H. V. Tran. Diel Patterns of Activity in Bacteria and Eukaryotes at the LEO-15 Study Site. Draft in prep for submission to FEMS Microbiol. Ecol.

Completed dissertations supported by the project:

1. 2006-2003. Darryl Babcock, M.S., Environmental Sciences. Thesis Title: Sequence Analysis of Large-insert BAC clones from Dominant Microorganisms in the Mid Atlantic Bight.
2. 2002-1999. Joshua Nelson, M.S., Oceanography, Thesis Title: Time series analysis using TRFLP analysis and DNA sequencing of 16S rDNA genes to assess background bacterioplankton variability and to identify dominant microorganisms off the coast of New Jersey.

Presented posters/papers supported by the project:

2003

1. Conference Presentation— Dissimilatory sulfite reductase genes (*dsrAB*) disclose the composition of suldifogenic communities in contaminated sediments worldwide. J. Perez-Jimenez, L. Young, and L. Kerkhof. ASM Gen. Meeting, Washington, DC, May 2003
2. Conference Presentation— Assessing active bacteria by fingerprinting ribosomes during a day/night cycle at LEO-15. L. J.Kerkhof, L. McGuinness, H. Tran ASLO Gen. Meeting, Salt Lake City, UT Feb 2003

2002

1. Conference Presentation (Invited)— **L. J. Kerkhof**, J. D. Nelson, H. V. Tran and C. Vetriani . High Density Sampling in the Coastal Ocean. ASLO Gen. Meeting, Honolulu, HI Feb 2002

2001

1. Conference Presentation-- C. Vetriani, H. V. Tran, and **L. J. Kerkhof** Phylogenetic and Functional Analysis of Microbial Communities at the oxic/anoxic interface in the Black Sea. ASM Gen. Meeting, Orlando, FL May 2001
2. Conference Presentation-- J. Perez-Jimenez, L. Young, and **L. Kerkhof**. Genetic Diversity of Sulfate Reducing Bacteria in a Contaminated and a Pristine Environment. ASM Gen. Meeting, Orlando, FL May 2001
3. Conference Presentation—J. D. Nelson and **L. J. Kerkhof** Monthly Time Series Analysis During 1995/1996 of Microbial Populations off the Coast of New Jersey. ASM Gen. Meeting, Orlando, FL May 2001

2000

1. Conference Presentation-- Perez-Jimenez, J., E. Sullivan, L. Young, and **L. Kerkhof**. Anaerobes and Biodegradation of Petroleum Components in Anoxic Sediment. American Society Microbiology (ASM), General Meeting, Los Angeles, CA, May.
2. Conference Presentation-- Vetriani, C. and **L. Kerkhof**. Phylogenetic and Functional Characterization of Microbial Assemblages at the Oxic/Anoxic Interface in the Black Sea. American Society of Limnology and Oceanography (ASLO). San Antonio, TX, January.
3. Conference Presentation-- Scala, D. and **L. Kerkhof**. Horizontal Heterogeneity in Denitrifying Bacterial Communities in Marine Sediments. American Society of Limnology and Oceanography (ASLO). San Antonio, TX, January.
4. Conference Presentation-- Rodriguez, A. and **L. Kerkhof**. Bacterial Community Fingerprints and Activity in a NearShore, Subsurface Jet at LEO-15 (a Long Term Ecosystem Observatory off the New Jersey Coastline.) American Society of Limnology and Oceanography (ASLO). San Antonio, TX, January.

Draft manuscript for submission to BMC Microbiology

Dominant Microorganisms in the Mid-Atlantic Bight

Lee Kerkhof[1*], Lora McGuiness[1], Eva Maslak[2], and Joshua Nelson[3]

[1] Institute for Marine and Coastal Sciences, 71 Dudley Road, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901

[2] Hospital of the University of Pennsylvania, Philadelphia, PA 19104

[3] Pt. Pleasant High School, Pt. Pleasant, NJ, 08742

Abstract:

A determination of dominant microorganisms in the Mid-Atlantic Bight was performed by terminal restriction length polymorphism analysis of 16S rRNA genes from monthly samples over a 2-year period in the Mid-Atlantic Bight (n=178). Frequency analysis of individual OTUs revealed 13 diagnostic T-RFLP peaks in over 66% of the samples collected during this time series. A subset of the SSU genes associated with these abundant TRFLP peaks were cloned and found to be nearly identical to sequences found in diverse geographic locations. Additionally, the cloned sequences were used to predict additional restriction enzyme cut sites for the specific genes to verify the results of the TRFLP frequency analysis. The results suggest that a small number of microbial species are the most abundant component of the coastal bacterioplankton in the MAB. These dominant bacteria are members of the alpha, gamma, and epsilon-Proteobacteria and likely important in biogeochemical processes in the coastal ocean.

Introduction:

It is well known that bacteria play a significant role in marine environments, utilizing 20-50% on average of primary production (8) and cycling much of the carbon, nitrogen, and sulfur in the oceanic biosphere (7). However, despite this importance of bacterial groups in marine systems, relatively little is known about the most abundant microbial species over wide temporal and spatial scales in the coastal ocean. Although many researchers have demonstrated specific biological, chemical, and physical factors that affect microbial populations in coastal, estuarine, and riverine environments (11, 15, 19, 23, 26), only recently, have suggestions been put forth for the dominant heterotrophic bacteria (Sar11 group) in open ocean, coastal, and polar environments (3, 24, 29). Most evidence presented to date claims dominance based on compiled

data from individual clone libraries or FISH probing of a relatively small number of samples. Other attempts to delineate the dominant microbial members of the marine community have focused on direct detection from a time series data set. For example, dominant denitrifying bacteria associated with coastal sediments were determined by frequency analysis of TRFLP profiles of nosZ genes collected over different time and space scales (Scala and Kerkhof, 2000). This approach was also used to analyze monthly surface samples from the San Pedro Channel during a 4 year period (Fuhrman et al, 2006). Finally,

In this study, we have attempted to assess the dominant bacteria in the Mid-Atlantic Bight (MAB) by performing frequency analysis on 178 small ribosomal subunit (SSU) TRFLP profiles collected along a 34 km transect over a two-year time period from 1995 to 1996 (27). This approach eliminates *E. coli* clonal bias which can preferentially enrich for alpha-Proteobacteria (10, 18) and does not require a large GC clamp during PCR for resolving different target genes. The results indicate that members of the *Roseobacter* clade and the epsilon *Proteobacteria* are part of the dominant bacteria in the coastal ocean off New Jersey. As the most abundant members of the bacterial community, these specific microbes likely play a major role in the cycling of carbon and nitrogen in the Mid-Atlantic Bight.

Materials and Methods

Water samples were collected once a month over a 3-year period (October 1994 to June 1997) along a 34 km transect off the New Jersey Coast as shown previously (27). Bacterial biomass was collected on 0.2 μm filters (500mL), frozen at sea, and stored at -80°C until DNA was extracted. One hundred seventy eight filters from 1995 and 1996 were analyzed for this study. Bacterial DNA was extracted using a previously described phenol/chloroform extraction with minor modifications (30). DNA was further purified in cesium chloride density gradients in

a Beckman XL-100 ultracentrifuge using a TLA 120.1 rotor. Purified DNA was visualized on a 1% agarose gel and quantified using a Fotodyne image analyzer (Fotodyne Inc., Hartland, WI) (16). 16S rRNA target genes were amplified using standard eubacterial primers 27F and 1525R, with 27F labeled with 6-FAM ((1)-carboxy-fluorescein) (Gibco BRL, Life Technologies, Gaithersburg, MD). Each 50 μL PCR reaction contained 0.2pM (each) primer, (dNTPs), 5uL of 10X Red Taq reaction buffer (Sigma-Genosys, Woodlands, TX), 1 U of Red-Taq polymerase (Sigma-Genosys), and between 5-10ng genomic DNA. Amplified DNA was visualized on a 1% agarose gel and quantified using a Fotodyne image analyzer using 250 ng of lambda *HinD* III size standard.

Twenty five ng of fluorescently labelled PCR products were digested with *MnlI* for primary TRFLP analysis and enzymes *AvaII*, *AluI*, *HaeIII*, and *HinPI* for subsequent analyses (New England Bio-Labs, Beverly, MA). Twenty μl digests were carried out at 37°C for 6 hours with 1U of restriction endonuclease in the manufacturer supplied reaction buffers. DNA was precipitated with 2 volumes of 95% ethanol followed by centrifugation at 14,000 rpm at 4°C in an Eppendorf microcentrifuge for 15 min. Pelleted DNA was rinsed twice with 250μL of 70% ethanol and dried in a vacuum centrifuge for 5 minutes. Fluorescently labeled DNA fragments were separated on an ABI Prism 310 autoanalyzer (Perkin Elmer Applied Biosystems) run in Genescan mode. Samples were resuspended in a solution containing 14.7μL deionized formamide and 0.3μL of TAMRA 500 bp size standard and denatured at 94°C for 3 minutes. Terminal restriction fragment (T-RF) size was determined by Genescan software version 2.1.1 (PE Applied Biosystems).

For cloning 16S rDNA genes, amplicons from 2 samples (5/96 A6B and 7/96 A6B) were added to a Topo TA cloning kit (Invitrogen Corporation, Carlsbad, CA) as per the manufacturer's

protocols. The resulting plasmids were transformed into chemically competent *Eschericia coli* (DH5α) on Luria agar plates containing 100μg/ml ampicillin. 200 clones were obtained from each library and screened by TRFLP using PCR amplified product directly from insert-containing colonies (17). Following PCR, three samples were pooled and digested with 1U of MnlI as described above. Samples that corresponded to the dominant fragments observed in the complex community T-RFLP profiles were purified using a Flexi-prep kit Pharmacia, Piscataway, NJ) and partial bidirectional sequence was obtained with the eubacterial primer 519R, and primers M13F and M13R on a Perkin-Elmer ABI 310 autoanalyzer (Perkin-Elmer, Foster City, CA).

Computer analyses of 16S rDNA sequence from clonal libraries was used to verify that sequence data contained the expected restriction site predicted in TRFLP screening of the clonal libraries using *MnlI*. The BLAST program was used to compare sequences obtained in this experiment with various sequences in the GenBank database. Initial sequence alignments, based on positions 27-477 (*E. coli* number system) were carried out using the program Sequence Navigator v. 1.0.1. Aligned sequences were also edited in the GDE program. The final unambiguous alignment was composed of 295 bases. Bootstrap values were calculated by the PHYLIP v. 3.6 program SEQBOOT and were based on 100 iterations. Phylogenetic trees were reconstructed using the PHYLIP v. 3.6 DNAML and CONSENSE (12). Accession numbers for the dominant SSU sequences are xxx-xxxx.

Results:

In order to assess whether TRFLP profiling could provide quantitative information on amplification products or target gene concentration in an environmental sample, a series of simple tests were performed. The first test examined the changes in peak area resulting from

increased amounts of amplicon added to the ABI 310 Gene Analyzer after precipitation and re-suspension. The results from triplicate samples of a 500 bp amplicon (0.5-12 ng) are shown in Fig. 1. A linear response for peak area versus mass is observed with tight clustering of most replicates. The second test involved adding genomic DNA from a laboratory culture to genomic DNA from salt marsh sediment. This test was designed to determine if a quantitative TRFLP response could be observed when a known mass of input DNA was added to a complex background of bacterial genomic DNA from the environment and subjected to PCR amplification, digestion, precipitation, and profiling. A logarithmic fit of the culture-specific peak area data demonstrates most variability in fluorescent signal is directly attributable to changes in input DNA (Fig. 2). This result is in good agreement with a prior report of highly reproducible profiles and peak areas for 86 biological replicates from 5 grey water treatment reactors (22).

These quantitative, reproducible TRFLP procedures were used for all samples in the MAB dataset. Specifically, an equal mass of genomic DNA was added to the initial PCR reaction to label SSU genes and an equal mass of amplicon was digested for the TRFLP profiling. This approach allowed us to identify the most abundant SSU genes in the MAB. A frequency plot of TRFLP peaks for the 178 profiles indicates that 13 SSU target genes dominate the fingerprints. These peaks are present in greater than 60% of the samples (Fig. 3). The major terminal restriction fragments (TRF's) were: 116, 125, 129, 131, 177, 205, 237, 241, 248, 250, 273, 276, and 282 bp. (The 177, 205, and 250 and 282bp TRFs were present in over 90% of the samples tested.) Thirty eight peaks were found to comprise the middle frequency region of the TRFLP histogram (from 33 % to 65 % frequency; (Fig. 3). The remaining 276TRF's were found to occur at the lowest frequencies, being present in less than 1/3 of the fingerprints, and likely represent those sequences commonly found in clone libraries or the rare biosphere.

The presence/absence data agrees well with a peak area analysis of the dataset. The results indicate the most frequently occurring TRF's also comprise the highest total peak area in the profiles for all 178 fingerprints (Fig 4) and represented the highest amounts of SSU target genes in the dataset. Among the various dominant TRF's, the 250bp peak had the highest overall area with an average relative proportion of 0.10 of the total peak area for the entire data set. These results suggest the 250 bp TRF phylotype is the most numerically abundant microbe in the MAB. TRF's 205 and 177 ranked second and third highest in total peak area and relative proportions, respectively. These high-frequency TRF's (250, 205, and 177 bp) accounted for ~62% of the overall total TRFLP area for the dataset, while the middle and lower frequency TRFs accounted for only 18% and 20% of total TRFLP area. The presence of a small group of frequently occurring TRFs with the largest TRF area in the dataset suggests that only a few microorganisms dominate the coastal ocean water column in the Mid-Atlantic Bight.

Through cloning and sequence analysis, six 16S rRNA genes were identified in 2 different clonal libraries from winter and summer samples that corresponded to 5 of the major TRFs in Figures 2a and 2b. These were the 129, 131, 177, 205, and 250 bp TRFs. (Two sequences were identified that had the endonuclease site for the 177bp MnlI restriction fragment.) Re-construction of phlyogenetic trees using the dominant TRF SSU genes with 16S rRNA sequences from GenBank revealed that the α, γ, and ε subdivisions of the *Proteobacteria* were recovered (Fig. 5).

Discussion:

Our quantitative TRFLP findings replicate earlier reports of correlations between TRFLP peak height and target gene abundance (9). Furthermore, the frequency analysis of multiple TRFLP profiles has been used previously to elucidate the dominant denitrifiers in coastal

sediments collected over time and space (30). The identification of the dominant heterotrophic bacteria in the Mid-Atlantic Bight using this approach is in close agreement with prior reports of group-specific data suggesting that α and γ-*Proteobacteria* are abundant members in the marine environment (14, 15, 25, 28, 29). The results also support the notion that clonal bias in *E. coli* recombinant libraries can significantly alter the proportion of recovered SSU clones (10, 18) since free living, aerobic ε-*Proteobacteria* are considered minor components of the microbial population for coastal environments except during phytoplankton blooms (19). In addition to affiliations with specific groups (e.g. α-*Proteobacteria*) many of the SSU clones that we describe are very closely related to phylotypes from diverse geographic locations, suggesting the dominant phylotypes from the MAB are globally distributed. For example, the 250 bp TRF clone is nearly identical (99.5%) to a 16S rRNA gene from Cape Hatteras, North Carolina (OM42) (26). One 177bp TRF clone was 99% similar to a bacterial isolate, HRV2 obtained from a marine sponge, *Halichondria panicea* in the Adriatic Sea (1). While another 177bp TRF clone shared 98.8% sequence identity with another Cape Hatteras clone OM55, (27). All these clones are within the α-*Proteobacteria*, with the 250 TRF related to the SAR83 cluster (27) which has been reported to contain bacteriochlorophyll *a* producing aerobic heterotrophs. If the 250 bp TRF is dominant and capable of anoxygenic photosynthesis, it is consistent with a recent finding indicating anoxygenic photo-aerobic heterotrophs comprise a significant proportion of photosynthetically active pigments in the MAB (21). Another SSU clone (131bp TRF) shared sequence identity (97%) to an unidentified ε-*Proteobacterial* sequence that was obtained from the Guaymas Basin off the coast of Mexico (32).

The remaining dominant SSU clones were unrelated to known SSU genes in Genbank. Specifically, the 129bp TRF clone was distantly related (93%) to a γ-*Proteobacterial* isolate from

Antarctic sea ice (6). While, the (205 bp TRF) clone was related (94%) to an uncultured γ-*Proteobacterial* sequence that was obtained from bacterial mats associated with black band disease on the surface of dead coral colonies (*Montastrea annularis)* in the Southern Caribbean Sea (13). Finally, a 235bp TRF clone , which represented a mid-range frequency of 56.5% of all profiles, was nearly identical (99% similarity) to a member of the SAR86 cluster collected off the coast of Cape Hatteras, NC (28). This clone is unique because its nearest relatives have been reported to use proteorhodopsin to carry out light-driven proton pumping and may convey a selective advantage (4, 5).

Although the dominant SSU genes from the MAB appear to have a global distribution, two potential methodological artifacts may account for the abundance of certain TRF's in the dataset. One possibility is that these particular SSU genes preferentially amplify from the extracts and overwhelm the other SSU genes present within the samples. However, a rank abundance curve of the TRFLP data indicates a smooth logarithmic decay of 327 OTUs with virtually no deviation from the curve (r^2=0.984; data not shown) suggesting there is little to no preferential amplification. In addition, a small number of fingerprints in the 2-year data set lack these dominant TRF peaks or the peaks are greatly reduced when compared to other SSU TRF peaks. These results indicate that preferential amplification is not occurring for the dominant TRFs discovered in this analysis. Another possibility for the predominance of certain TRFs is that multiple SSU genes underlie any particular peak in the fingerprint. For example, the 177 bp TRF was found in 2 different clones within our libraries. It is possible that the 250 or 205 TRF are also comprised of multiple SSU genes, resulting in the appearance of dominance in the dataset. To test the notion of multiple SSU genes underlying dominant TRF peaks, additional restriction digests were conducted on additional samples as shown in Fig. 6. This schematic demonstrates

how a single peak for restriction enzyme A (containing 2 OTUs) can be resolved by additional enzyme digests using restriction enzyme B. If the peak area for the predicted TRF is lower than the original TRFLP peak (by a factor of 2 in this case), there must be more than one OTU representing the original TRFLP peak. For the MAB sample set, we tested whether a single SSU gene accounted for the various dominant peaks in the environmental fingerprints by digesting with enzymes that cut in the cloned genes we had obtained from samples representing seasonal groups in the MAB (27). If there were only a single OTU in the dominant TRFLP peaks, then digests with additional restriction enzymes would lead to predictable TRF peak areas within the profiles from other samples that are of a comparable peak area to the original *Mnl* I digest. The results using eight different samples from 2/96, 5/96, 7/96 and 10/96 with the restriction endonucleases *AluI* and *AvaII* are summarized in Table 1. Observed peak areas for *Mnl* I and the expected TRFs for the additional secondary enzymes agreed well for the 129, 131, and 205 bp dominant peaks, except where no TRF fragment of the expected size was detected for a particular enzyme. However, in many of these secondary digests lacking the predicted TRF peak, a diagnostic peak representing the 2nd cut site with respect to the fluorescent label for the particular enzyme is detected. This result implies some polymorphisms with *Alu* I and *Ava* II cut site and that 1-2 phylotypes can account for the dominant TRF's seen in the overall dataset. In contrast, the agreement between peak area for the additional digests of the 177 bp amplicon were not as robust. The percentage of the initial *Mnl* I peak area that was detectable in the secondary digests was found to be much less than for the other dominant TRF's indicating additional phylotypes underlie this particular *Mnl* I peak, as exemplified by the 2 different clones which share a 177 cut site with *Mnl* I. Whether the differences between predicted and observed TRF's peak areas results from different SSU genes sharing many terminal restriction sites or sequence

microheterogeneity in copies of SSU operons within a particular genome cannot be completely determined with this dataset. However, it is possible to calculate the possibility of 2 OTUs sharing any particular TRF by determining the probability of not cutting (i.e. 255/256 for a 4 bp restriction enzyme) raised to the number of bases between the primer and the terminal restriction site times 1/256 for the recognition site. This is roughly 10^{-3} for any particular enzyme. Therefore, the probability that 2 OTUs share 3 TRFs is roughly 10^{-9} for any randomly generated sequence.

In conclusion, our results demonstrate that a small group of high-frequency, phylotypes can account for over 60% of the total TRFLP area in the Mid Atlantic Bight. These findings are in agreement with reports that members of the α-subclass of *Proteobacteria* are significant and numerically abundant in the coastal ocean (3, 14, 25, 31). Yet, our results also suggest that it is not the entire group of α-*Proteobacteria* that are in high numbers in the coastal ocean; but, only a few principle members of this subdivision. Additionally, our data contrast with reports of the SAR11 group, as a dominant bacterioplankton lineage (24). A computer-simulated endonuclease digestion of several SAR11 sequences from Genbank revealed that the SAR 11 TRF's in our MAB dataset are present either at very low frequencies (<2%), or are absent from the bacterial community in our 2-year study. These results imply that many known members of the SAR11 clade play a minor role in the Mid Atlantic Bight. However, it may be that unknown members of the SAR 11 group comprise the remaining 7 TRF's with high frequency that have yet to be identified. Finally, our data demonstrate it is possible to move beyond a group specific approach to microbial ecology in the ocean. Efforts can now focus on the dominant microorganisms within the bacterial community to improve our knowledge about their autecology and gain

further insights into their metabolism and their effects on biogeochemical cycling in the coastal ocean.

Acknowledgements:

This research was supported by funds from NOAA-NURP through the Mid-Atlantic Bight Center, the National Science Foundation (OCE-9872024), and DOE Biotechnological Investigations – Ocean Margin Program (BIOMP; DEFG02- 00ER62978) to L.J.K. We also wish to thank Sue Boehme, Clare Reimers, Elizabeth Creed, Paul Field, David Scala, Mary Voytek, Rob Sherrell, Rick Laubly, John Zlotnik and Rose Petrecca for their assistance with sample collection.

Figure Legends:

Figure 1: TRFLP fluorescent peak area vs. mass of a 500 bp amplicon (ng)

Figure 2: TRFLP fluorescent peak area vs mass of input DNA for a specific peak in a background of complex environmental DNA

Figure 3: TRF frequency for all fingerprints generated in the Mid Atlantic Bight (n= 178 samples). The peaks present at frequencies of >66% are indicated.

Figure 4: TRF total peak area for all fingerprints generated in the Mid Atlantic Bight (n= 178 samples). The peaks with the 10 highest areas are indicated.

Figure 5: Maximum likelihood tree representing the 16S rRNA genes that correspond to the dominant TRFs using 295 unambiguously aligned bases. Bootstrap values ≥50 are indicated.

Figure 6: Schematic demonstrating the principle of using secondary digests and peak area information to discern if a TRFLP is composed of more than one OTU.

References

1. **Althoff, K., C. Schütt, R. Steffen, R. Batel, and W. E. G. Müller.** 1998. Evidence for a symbiosis between bacteria of the genus Rhodobacter and the marine sponge Halichondria panicea: harbor also for putatively toxic bacteria? Marine Biology **130**:529-536.
2. **Avaniss-Aghajani, E., K. Jones, D. Chapman, and C. Brunk**. 1994. A molecular technique for identification of bacteria using small subunit ribosomal RNA sequences. Biotechniques **17**:144-6.
3. **Bano, N., and J. T. Hollibaugh**. 2002. Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. Applied and Environmental Microbiology **68**:505-518.
4. **Béjá, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. DeLong**. 2000. Bacterial rhodopsin: evidence for a new type of phototrohpy in the sea. Science **289**:1902-1906.
5. **Béjá, O., M. T. Suzuki, J. F. Heidelberg, W. C. Nelson, C. M. Preston, T. Hamada, J. A. Elsen, F. C. M., and E. F. DeLong**. 2002. Unsuspected diversity among marine aerobic anaoygenenic phototrophs. Nature 415.
6. **Bowman, J. P., S. A. McCammon, M. V. Brown, D. S. Nichols, and T. A. McMeekin**. 1997. Diversity and association of psychrophilic bacteria in antarctic sea ice. Applied & Environmental Microbiology **63**:3068-3078.
7. **Carpenter, E**. 1983. Nitrogen in the marine environment. Academic Press, London, UK.
8. **Church, M. J., D. A. Hutchins, and H. W. Ducklow**. 2000. Limitation of bacterial growth by dissolved organic matter and iron in the Southern ocean. Appl. Environ. Microbiol. **66**:455-66.
9. **Clement, B. G., and C. L. Kitts**. 2000. Isolating PCR-quality DNA from human feces with a soil DNA kit. Biotechniques **28**:640-+.
10. **Cottrell, M. T., and D. L. Kirchman**. 2000. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. Appl Environ Microbiol **66**:5116-22.
11. **Crump, B. C., E. V. Armbrust, and J. A. Baross**. 1999. Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. Appl. Environ. Microbiol. **65**:3192-3204.
12. **Felsenstein, J**. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics **5**:164-166.
13. **Frias-Lopez, J., A. L. Zerkle, G. T. Bonheyo, and B. W. Fouke**. 2002. Partitioning of bacterial communities between seawater and healthy black band diseased and dead coral surfaces. Appl. Environ. Microbiol. **68**:2214-2228.
14. **Gonzalez, J. M., and M. A. Moran**. 1997. Numerical dominance of a group of marine bacteria in the alpha-subclass of the class proteobacteria in coastal seawater. Appl. Environ. Microbiol. **63**:4237-4242.
15. **González, J. M., R. Simó, R. Massana, J. S. Covert, E. O. Casamayor, C. Pedrós-Alió, and M. A. Moran**. 2000. Bacterial community structure associated with a

dimethylsulfoniopropionate- producing North Atlantic algal bloom. Appl. Environ. Microbiol. **66**:4237-4246.

16. **Kerkhof, L**. 1997. Quantification of total RNA by ethidium bromide fluorescence may not accurately reflect the RNA mass. Journal of Biochemical & Biophysical Methods **34**:147-154.
17. **Kerkhof, L., M. Santoro, and J. Garland**. 2000. Response of soybean rhizosphere communities to human hygiene water addition as determined by community level physiological profiling (CLPP) and terminal restriction fragment length polymorphism (TRFLP) analysis. Fems Microbiology Letters **184**:95-101.
18. **Kerkhof, L., and D. Scala**. 2000. Molecular Tools for Studying Biogeochemical Cycling in Salt Marshes, p. 443-469. In M. Weinstein and D. Kreeger (ed.), Concepts and Controversies in Tidal Marsh Ecology. Kluwer Academic, Dordrecht/Boston/London.
19. **Kerkhof, L. J., M. A. Voytek, R. M. Sherrell, D. Millie, and O. Schofield**. 1999. Variability in bacterial community structure during upwelling in the coastal ocean. Hydrobiologia **401**:139-148.
20. **Kitts, C. L**. 2001. Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics. Curr Issues Intest Microbiol **2**:17-25.
21. **Kolber, Z. S., C. L. Van Dover, R. A. Niederman, and P. G. Falkowski**. 2000.Bacterial photosynthesis in surface waters of the open ocean. Nature **407**:177-9.
22. **McGuinness, L. M, Salganik, M., Vega, L., Pickering, K. D. and Lee J. Kerkhof**. 2006. Replicability of Bacterial Communities in Denitrifying Bioreactors as Measured by PCR/T-RFLP Analysis. Env. Science and Tech. **40**: 509-515.
23. **Moeseneder, M. M., C. Winter, J. M. Arrieta, and G. J. Herndl**. 2001. Terminal-restriction fragment length polymorphism (T-RFLP) screening of a marine archaeal clone library to determine the different phylotypes. Journal of Microbiological Methods **44**:159-172.
24. **Morris, R. M., M. S. Rappe, S. A. Connon, K. L. Vergin, W. A. Siebold, C. A. Carlson, and S. J. Giovannoni**. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature **420**:806-810.
25. **Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni**. 1995. Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnology and Oceanography **40**:148-158.
26. **Murray, A. E., J. T. Hollibaugh, and C. Orrego**. 1996. Phylogenetic compositions of bacterioplankton from two california estuaries compared by denaturing gradient gel electrophoresis of 16S rdna fragments. Appl. Environ. Microbiol. **62**:2676-2680.
27. **Nelson, J. D., Boehme, S. E., Reimers, C. E., Sherrell, R. M., and L. J. Kerkhof**. 2008. Temporal patterns of microbial community structure in the Mid-Atlantic Bight. FEMS Micro. Ecol. **65**: 484-493
28. **Rappé, M. S., P. F. Kemp, and S. J. Giovannoni**. 1997. Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. Limnology and Oceanography **42**:811-826.
29. **Rappé, M. S., K. Vergin, and S. J. Giovannoni**. 2000. Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. FEMS Microbiol Ecol. **33**:219-232.
30. **Scala, D. J., and L. J. Kerkhof**. 1999. Diversity of nitrous oxide reductase (*nosZ*) genes in continental shelf sediments. Appl. Environ. Microbiol. **65**:1681-1687.

31. **Suzuki, M. T., M. S. Rappe, Z. W. Haimberger, H. Winfield, N. Adair, J. Stroebel, and S. J. Giovannoni**. 1997. Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Appl. Environ. Microbiol. **63**:983-989.
32. **Teske, A., K.-U. Hinrichs, V. Edgcomb, A. de Vera Gomez, D. Kysela, S. P. Sylva, M. L. Sogin, and H. W. Jannasch**. 2002. Microbial Diversity of Hydrothermal Sediments in the Guaymas Basin: Evidence for Anaerobic Methanotrophic Communities. Appl. Environ. Microbiol. **68**:1994-2007.

Draft manuscript for submission to Microbial Ecology

Analysis of Two 30kb DNA Fragments From Dominant Bacteria in the Mid-Atlantic Bight

Darryl A. Babcock[1], B. Wawrik[2], J. H. Paul[3], L. Mcguinness[1], L. J. Kerkhof[1*]

AFFILIATION:
[1]Institute of Marine and Coastal Sciences, Rutgers University;
[2]Department of Botany and Microbiology, University of Oklahoma;
[3]College of Marine Science, University of South Florida

Abstract:

A bacterial artificial chromosome (BAC) library was generated using community DNA extracted from the LEO-15 site off the New Jersey coast and screened for dominant 16S rRNA genes identified in prior studies. Two clones belonging to the *Roseobacter* CHAB-I-5 clade, believed to be significant components of the microbial population in many marine environments, were identified and sequenced. These *Roseobacter* BACs were annotated for protein coding sequences, as well as comprehensive G+C content analysis and phylogenetic testing. The analysis also revealed that both *Roseobacter* BACs contained antibiotic resistance genes: a beta-lactamase gene and a RND drug exporter. Additionally, the BAC's contained evidence of significant lateral gene transfer (LGT) with respect to the other *Roseobacter* populations. This antibiotic resistance and enhanced LGT may contribute to the dominance of these particular microorganisms in the marine environment in the Mid-Atlantic Bight, surrounding the LEO-15 study site.

Introduction

There is little understanding of *dominant* microorganisms in the ocean and few insights into the mechanisms leading to the success of one bacteria over the other in environmental systems. Previous studies along a 40 km transect off the Jersey Coast identified molecular signals from dominant bacteria within the microbial community by assessing terminal restriction fragment length polymorphism TRFLP; 1) peak frequency over a 2-year time period (n=178; 16; Kerkhof et al., companion paper submitted). A total of 13 dominant microbial signatures were identified in this TRFLP analysis which were present in over 60% of all onshore/offshore profiles throughout the time series. The rationale behind this current study was to identify clones matching these dominant bacteria in a large-insert, bacterial artificial chromosome library (BAC) from the Mid Atlantic Bight to glean additional insights into their ecology. The metagenomic approach has, among other studies, been used to analyze segments of the genomes of an uncultivated archaeon (22), to discover a new method of phototrophy in marine prokaryotes (3), to analyze the community genomics of bacteria within the oceans interior (6), and to identify genome organization surrounding the Rubisco gene in marine bacterioplankton. (11)

For this study, two BAC clones containing 16S genes which produce TRFLP peaks that correspond to dominant peaks in the Mid-Atlantic Bight (Kerkhof et al., companion paper submitted) were identified in a BAC library created from summer samples in 2002 using a rapid, multi-screening approach based on TRFLP methods (2). These BACs (designated p248 and p250 corresponding to the TRFLP peaks resulting from *Mnl I* digestion) were purified, sequenced, and annotated. The genomic DNA belonged to two members of the *Roseobacter* clade. Each pBAC was approximately 30 kb in size and contained a complete ribosomal operon, multiple protein coding sequences, including genes involved in antibiotic resistance.

Furthermore, analysis of the coding regions demonstrated lateral gene transfer of multiple proteins and provided an estimate of entire genome G+C content for the microorganisms originating the pBACs. These results suggest major differences in overall G+C content among dominant members of the *Roseobacter* clade, implying enhanced lateral gene transfer. This lateral gene transfer (LGT) may provide the means by which the p248 and p250 microorganisms adapt and becomes dominant within their local microbial community.

Materials and method:

Extraction and sequencing of BAC clones:

The BAC library was created using the pIndigoBac system (Epicenter, Madison, WI) and screened by TRFLP as described previously (2). After purification, the p248 and p250 clones were grown in 50ml volumes with a chloramphenicol concentration of 100 mg L^{-1} at 37°C for approximately 15 hours. The BAC MAX DNA Purification kit was used for DNA purification (Epicentre biotechnologies, Madison, WI). BACs were digested using *Not*I (New England Biolab, Beverly, MA) to remove vector from the inserted DNA. The inserted DNA was gel purified using the Geneclean II Kit (Qbiogene, Solon, OH) and mechanically sheared to 1-2 Kb with a DNA Hydroshear (Genomic Solutions; Ann Arbor, MI). The DNA ends was repaired using an End-It™ Kit (Epicentre, Madison, WI) and blunt-end ligated into pBluescript. The ligation product was transformed into electo-competent cells and the resulting colonies were subjected to blue/white screening using X-gal and IPTG. Recombinant plasmids, with inserts estimated to be greater than 500 bps, were purified using a FlexiPrep Kit (Amersham Biosciences, Piscataway, NJ) and bidirectionally sequenced with M13forward and reverse primers on a Perkin-Elmer ABI 310 autoanalyzer (Perkin-Elmer, Foster City, CA). Gaps were

filled in by primer walking. The various DNA sequences were assembled using the Auto Assembler program. (Perkin-Elmer/ABI).

Annotation and metagenomic analysis of BACs:

Open reading frames were detected using ORF finder, BLASTX , and the sequence alignment program on the NCBI webpage (http://www.ncbi.nlm.nih.gov/). Comparative genomic analysis was conducted using the JGI webpage (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi). G+C plot analysis was performed using the Artemis bioinformatics computer program. (21). Regions displaying a rise or decline in the G+C plot when compared with the estimate overall G+C content were subjected to phylogenetic analysis to assess lateral gene transfer events. Phylogenetic trees were re-constructed using Mega2: genetic analysis software. (13) with neighbor joining methods and boot strap analysis (n=500). Overall genomic G+C content for the p248 and p250 BACs was estimated by assessing the mean G+C content of 10 protein-coding genes. (24) and the estimate was tested against 7 *Roseobacter* genomes in the JGI data base (Table 1).

Results:

Gene organization within the pBAC's.

Analysis of the phylogenetic relationships for SSU genes from pBAC 248 and pBAC 250 indicated that the microorganisms associated with these dominant TRFLP peaks were members of the α-Proteobacteria with closest relatives in the *Roseobacter* group (Fig. 1). The pBAC sequences clustered with the CHAB-I-5 group of the *Roseobacteria*, comprised solely of environmental sequences from coastal environments (5). The insert in pBAC 248 was approximately 32 Kb, with 20 open reading frames and an entire rRNA gene operon (Fig. 2). Homology searches of these ORFs suggested pBAC 248 contained an operon for ABC dipeptide

transporters, a metallo-β-lactamase gene, and genes involved in the oxidation of sulfite to sulfate (Supplemental Material 1). Over half of these ORF's (13 of 20) are most closely related to other members of the *Roseobacter* clade, according to BLAST searches. The remaining 7 ORF's correspond to probable proteins that cluster around an integrative genetic element/ resolvase and share homologies with other *Proteobacteria*, *Firmicutes*, or *Cyanobacteria*.

The insert from pBAC 250 was approximately 27 Kb, consisting of 16 open reading frames and the entire rRNA gene operon. The majority of the coding regions were involved in the synthesis of lipopolysaccarides while 3 genes were putatively identified as a probable RND multidrug exporter (Fig. 2). Seven probable ORFs have a strong homology with proteins found in other members of the *Roseobacter* clade. The remaining ORF's are clustered around two sections with protein homologies associated with α-*Proteobacteria*, *Cyanobacteria*, *Acidobacteria*, and Archaea. The second cluster has the closest protein affiliations with the γ-*Proteobacteria* (Supplemental material #2).

The genomic data collected from pBAC 248 reveals two syntenic regions within the BACs gene organization. The first syntenic region contains ABC transporters and is highly conserved within *Roseobacter* clade (Fig. 3). The second syntenic region includes the rRNA operon and a ferric reductase-like gene which may be involved in electron transport, an oxidoreductase molybdopterin binding domain protein which may catalyse the oxidation of sulphite to sulphate, and an AAA ATPase involved in the assembly, operation, or disassembly of protein complexes. In contrast, there is little conserved gene synteny observed in the genes on pBAC 250, excluding the ribosomal operon (Fig. 3). Of the seven protein encoding genes which show evidence of being descended vertically through the *Roseobacter* clade only 4 are found in

more than 50% of a set of *Roseobacter* reference genomes used in this study (Table 1) and only 2 of these genes are found in all of the reference genomes.

Detection of Laterally transferred genes on the BACs

This high percentage of coding regions, with BLAST homologies different than the *Roseobacter* clade, suggested that many genes within the pBACs may have resulted from lateral gene transfer. In order to verify if these open reading frames were not related to *Roseobacter* genes, a search for a significant rise or decline in the overall G+C content for the entire pBAC was initiated (Fig. 2). This approach indicates most genes on pBAC 248 tracks with overall G+C content, except for 1 minor rise between 4-6 Kb and two significant depressions between 9 -12 Kb and between 17-21.5 Kb. The 4.3-6.3 Kb segment encodes a putative 2-dehydropantoate 2-reductase gene, which is most similar to a β-Proteobacterium by BLAST search. The section between 9-12 Kb contains a putative tetratricopeptide (TPR) repeat, a structural element of proteins thought to contribute to protein-protein interaction or protein stability related to the *Bacteriodetes*. Finally, the 17-21.5 Kb region contains six ORFs with varied phylogenetic ancestries, including a metallo β-lactamase-like gene by BLAST.

The G+C plot for pBAC 250 is slightly more uniform than the G+C plot for pBAC 248. There is a minor depression at the beginning of the insert associated with a putative mannuronan C-5-epimerase which is loosely conserved within the *Roseobacter* clade. The remaining regions deviating from overall G+C content correspond to a 3 ORF cluster, containing a (RND) drug exporter, an ion transport protein. Both coding regions are predominantly homologous to γ-Proteobacteria by BLAST. Re-construction of the phylogenetic relationship with this putative protein confirms the metallo β-lactamase-like gene on pBAC 248 is very distantly related to homologous genes found within the *Roseobacter* Clade (Fig. 4). Interestingly, the pBAC 248

gene is equally distant with many different bacterial phyla, including *Proteobacteria*, *Chloroflexi*, *Firmicutes*, and *Cyanobacteria*. Likewise, the RND exporter from pBAC 250 is more closely related to γ-Proteobacteria rather than *Roseobacters,* including *Photobacterium profundum* and *Saccarophagus degradens*.

Guanine and Cytosine (G+C) content of p248 and p250 BACs

The lack of synteny and significantly different homology of portions of the different pBACs to other bacteria in Genbank suggested that significant lateral gene transfer (LGT) may have occurred in the respective genomes. However, without the complete genomic sequence for the pBAC 248 and pBAC 250 *Roseobacters*, it is difficult to test this hypothesis. To address this issue, we estimated the overall G+C content for the entire genomes of the pBAC 248 and 250 microorganisms by averaging the G+C content of unbiased coding genes within the inserts (24). (A biased gene was considered to be a gene that clearly did not descend vertically from the *Roseobacter* clade and would not accurately reflect the evolving G+C usage within that bacteria's genome.) Using these criteria 12 genes were selected to be analyzed on pBAC 248 (Table 2) and 7 genes on pBAC 250 (Table 3). These genes predicted the entire genome G+C contents for the microorganisms represented by pBAC 248 and pBAC 250 to be 51.5% and 51.8% respectively. The approach was confirmed by testing those same unbiased genes to predict the genomic G+C content of 7 reference *Roseobacters* genomes (Table 1), which have been completely sequenced. Tables 2 and 3 summarize the results of this method applied to pBACs 248 and 250 as well as the method applied to the reference genomes. Most of the estimated predictions of the reference *Roseobacter* genomes are within a 2% deviation of the actual G+C content of their respective genomes. The G+C estimates using the gene sequences from pBAC 248 systematically underestimated the actual G+C content, while the sequences from pBAC 250

overestimated the G+C content. Only 2 whole genome estimates were 3-4% different from the actual G+C content of the reference microorganisms, reflecting the small number of genes on pBAC 250 that were present in these particular reference *Roseobacters*. These results suggest an overall 8-16% difference in G+C content for the microorganisms originating pBAC 248 and 250 with respect to other *Roseobacters*. The most likely way to achieve such stark differences in overall G+C content while maintaining a strong phylogenetic relationship (Fig. 1) is to have enhanced lateral gene transfer.

Discussion:

The sequences from the pBACs at LEO-15 provide some intriguing clues as to the mechanisms employed to achieve dominance by certain *Roseobacters* in the Mid-Atlantic Bight. Although there will be many factors that contribute to success of a given bacterium within any particular environment, enhanced antibiotic resistance is an example of a genotype which may provide a selective advantage for pBAC 248 and 250. Both of the pBACs investigated during this study contained antibiotic resistance genes. Specifically, pBAC 248 harbored a metallo-β-lactamase gene. This class of lactamases has only been discovered within the past 25 years, primarily because they are not usually found within pathogenic bacteria (20). The metallo-lactamases are extremely efficient at hydrolyzing most of the known families of beta-lactams (23), which may be produced by marine fungi and actinomycetes (4, 14). It is therefore conceivable that a beta-lactamase gene may confer a competitive advantage to the pBAC 248 bacterium, particularly in micro-environments, such as particles, where competition may be strong.

Likewise, the pBAC 250 contains a gene cassette of a predicted RND exporter gene. RND exporters provide defense against a wide range of lipophilic antibiotics, various dyes, and

detergents that can be toxic to bacteria. A RND operon usually consists of four genes: a transporter gene, a membrane fusion protein (MFP), an outer membrane protein (OMP), and a repressor gene (17). The RND gene cassette on pBAC 250 lacks the repressor gene, which could lead to constituative epression of the RND exporter. Studies conducted on an efflux system of *Neisseria gonorrhoeae* reveal that deletions of the *mtr*R repressor gene lead to over expression of *mtr*C and increased antibiotic resistence. (9) Similar studies conducted on the AcrAB and AcrEF efflux systems revealed that mutations in one of three regulators, AcrR, MarR, or SoxS, lead to increased expression of the AcrAB efflux system (10). Although it may be possible that a repressor gene may be located elsewhere in the genome, it would be interesting to test whether these dominant bacteria are highly resistant to bacterial antiobiotics.

The role of lateral gene transfer and its effect on the ecology of microbial populations in the environment is not yet completely understood, but is conceivable that bacteria can gain a competitive advantage through the acquisition of new genes by lateral gene transfer (18). This study indicates that the *Roseobacter* strains, originating the pBAC 248 and 250 sequences, may have a greater than expected proportion of genes originated from lateral transfer. Each of the pBACs contains genes for transport proteins, with some genes that are associated with other bacterial or archaeal phyla and some that are part of the α-Proteobacteria. One such example is the conserved genes within pBAC 248 that exhibit high percentage identity to a family of ABC transporters involved in the uptake of dipeptides/ oligopeptides/ and nickel. These types of transporters are found in many species of prokaryotes and are important in transporting nutrients into the cell. However, there are reports that another *Roseobacter* (*Silicibacter)* also contains extraordinarily large variety of transporters for various compounds ranging from amino acids, putrescine, DMSP, ammonium, to taurine in its genome which could lead to a selective

advantage for this group of bacteria. (15) Since coastal ecosystems can be subject to sudden changes from algal blooms or an influx of carbon from terrestrial environments, a high number of transporters, numerous metabolic pathways, and multiple copies of the rRNA operon would allow these *Roseobacters* to take advantage of a sudden increase of a variety of nutrient sources (15).

Finally, recent literature has revealed the importance of *Roseobacter* clade as an active component of the cycling of sulfur within marine ecosystems. The research has focused on the cycling of dimethylsulfoniopropionate (DMSP), produced by marine phytoplankton as an osmotic protectant. On method of degradation involves the production of dimethyl sulfide (DMS) while another pathway produces methanethiol (MESH). The MeSH can be incorporated into the amino acids of the bacteria or used as an energy source, by oxidizing HS^- to SO_4^{2-}. (12) Recently, it was demonstrated that both *Roseovarious Nubinhibens ISM* and *Silicibacter Pomeryi* where capable of these dual DMSP degradation pathways. (7) Considering that *Roseobacter*s have been associated with algal blooms (8, 19), it is possible that DMSP could be an important source of both carbon and energy for this group in the Mid-Atlantic Bight. For example, on pBAC 248, genes were discovered that are homologous to a sulfite oxidase which can create a proton motive force leading to ATP production coupled to the oxidation of sulfite to sulphate.

In conclusion, metagenomic analysis was used to identify and sequence two BAC clones containing ribosomal operons of dominant *Roseobacter* species in the Mid Atlantic Bight. Many genes contained on the pBACs 248 and 250 clones may contribute to the dominance of these bacteria within the coastal marine environment. Evidence is presented of enhanced lateral gene transfer of multiple coding sequences involved in transport and antibiotic resistance. These data provide a new range of testable hypotheses for bacterial ecology in the coastal ocean and

promise to be valuable in unveiling the mechanisms by which bacteria interact and dominate within the microbial community.

Acknowledgments:

This research was supported in part by grants from of the U.S. Department of Energy Biotechnological Investigations—Ocean Margin Program (BIOMP; DE-FG02-00ER62978 and DE-FG02-06ER64149) and the National Science Foundation (NSF; OCE 98-72024) to L.J.K. We thank Hiep Tran and Lora McGuinness for sample collection and logistics and the crew of the RVs Arabella and Caletta for sample collection and transport.

References:

1. **Avaniss-Aghajani, E., K. Jones, D. Chapman, and C. Brunk**. 1994. A molecular technique for identification of bacteria using small subunit ribosomal RNA sequences. Biotechniques **17**:144-6.
2. **Babcock, D. A., B. Wawrik, J. H. Paul, L. McGuinness, and L. J. Kerkhof**. 2007. Rapid screening of a large insert BAC library for specific 16S rRNA genes using TRFLP. Journal of Microbiological Methods **71**:156-161.
3. **Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. DeLong**. 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. Science **289**:1902-1906.
4. **Bhadury, P., B. T. Mohammad, and P. C. Wright**. 2006. The current status of natural products from marine fungi and their potential as anti-infective agents. Journal of Industrial Microbiology & Biotechnology **33**:325-337.
5. **Buchan, A., J. M. Gonzalez, and M. A. Moran**. 2005. Overview of the marine *Roseobacter* lineage. Applied and Environmental Microbiology **71**:5665-5677.
6. **DeLong, E. F., C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N. U. Frigaard, A. Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm, and D. M. Karl**. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science **311**:496-503.
7. **Gonzalez, J. M., J. S. Covert, W. B. Whitman, J. R. Henriksen, F. Mayer, B. Scharf, R. Schmitt, A. Buchan, J. A. Fuhrman, R. P. Kiene, and M. A. Moran**. 2003. *Silicibacter pomeroyi* sp nov and *Roseovarius nubinhibens* sp nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. International Journal of Systematic and Evolutionary Microbiology **53**:1261-1269.
8. **Gonzalez, J. M., R. Simo, R. Massana, J. S. Covert, E. O. Casamayor, C. Pedros-Alio, and M. A. Moran**. 2000. Bacterial community structure associated with a

dimethylsulfoniopropionate-producing North Atlantic algal bloom. Applied and Environmental Microbiology **66**:4237-4246.

9. **Hagman, K. E., and W. M. Shafer**. 1995. Transcriptional Control of the Mtr Efflux System of Neisseria-Gonorrhoeae. Journal of Bacteriology **177**:4162-4165.
10. **Jellen-Ritter, A. S., and W. V. Kern**. 2001. Enhanced expression of the multidrug efflux pumps AcrAB and AcrEF associated with insertion element transposition in Escherichia coli mutants selected with a fluoroquinolone. Antimicrobial Agents and Chemotherapy **45**:1467-1472.
11. **John, D. E., B. Wawrik, F. R. Tabita, and J. H. Paul**. 2006. Gene diversity and organization in rbcL-containing genome fragments from uncultivated Synechococcus in the Gulf of Mexico. Marine Ecology-Progress Series **316**:23-33.
12. **Kiene, R. P., L. J. Linn, and J. A. Bruton**. 2000. New and important roles for DMSP in marine microbial communities. Journal of Sea Research **43**:209-224.
13. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei**. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244-1245.
14. **Lam, K. S**. 2006. Discovery of novel metabolites from marine actinomycetes. Current Opinion in Microbiology **9**:245-251.
15. **Moran, M. A., A. Buchan, J. M. Gonzalez, J. F. Heidelberg, W. B. Whitman, R. P. Kiene, J. R. Henriksen, G. M. King, R. Belas, C. Fuqua, L. Brinkac, M. Lewis, S. Johri, B. Weaver, G. Pai, J. A. Eisen, E. Rahe, W. M. Sheldon, W. Y. Ye, T. R. Miller, J. Carlton, D. A. Rasko, I. T. Paulsen, Q. H. Ren, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, M. J. Rosovitz, D. H. Haft, J. Selengut, and N. Ward**. 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature **432**:910-913.
16. **Nelson, J. D., S. E. Boehme, C. E. Reimers, R. M. Sherrell, and L. J. Kerkhof**. 2008. Temporal patterns of microbial community structure in the Mid-Atlantic Bight. FEMS Microbiology Ecology **65**:484-493.
17. **Nikaido, H**. 1996. Multidrug efflux pumps of gram-negative bacteria. Journal of Bacteriology **178**:5853-5859.
18. **Ochman, H., J. G. Lawrence, and E. A. Groisman**. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405**:299-304.
19. **Pinhassi, J., R. Simo, J. M. Gonzalez, M. Vila, L. Alonso-Saez, R. P. Kiene, M. A. Moran, and C. Pedros-Alio**. 2005. Dimethylsulfoniopropionate turnover is linked to the composition and dynamics of the bacterioplankton assemblage during a microcosm phytoplankton bloom. Applied and Environmental Microbiology **71**:7650-7660.
20. **Rossolini, G. M., M. A. Condemi, F. Pantanella, J. D. Docquier, G. Amicosante, and M. C. Thaller**. 2001. Metallo-beta-lactamase producers in environmental microbiota: New molecular class B enzyme in *Janthinobacterium lividum*. Antimicrobial Agents and Chemotherapy **45**:837-844.
21. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell**. 2000. Artemis: sequence visualization and annotation. Bioinformatics **16**:944-945.
22. **Stein, J. L., T. L. Marsh, K. Y. Wu, H. Shizuya, and E. F. DeLong**. 1996. Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment front a planktonic marine archaeon. Journal of Bacteriology **178**:591-599.

23. **Walsh, T. R., M. A. Toleman, L. Poirel, and P. Nordmann**. 2005. Metallo-beta-lactamases: the quiet before the storm? Clinical Microbiology Reviews **18**:306-325.
24. **Zavala, A., H. Naya, H. Romero, V. Sabbia, R. Piovani, and H. Musto**. 2005. Genomic GC content prediction in prokaryotes from a sample of genes. Gene **357**:137-143.

Table 1: Protein coding sequences on PBAC 248.

Gene Name	ORF/ Dir.	# of AA	Species Affinity	% gene homology	Predicted gene function
Peptide ABC Transporter	1220 to 3; -	406	α-*proteobacteria Roseobacter*	63%/ 77%	ATP-binding protein
Peptide ABC Transporter	2130 to 1210; -	306	α-*proteobacteria Roseobacter*	64%/ 80%	membrane spanning protein (oligopeptide)
Peptide ABC Transporter	3068 to 2127; -	313	α-*proteobacteria Roseobacter*	64%/ 77%	permease protein
Peptide ABC Transporter	4766 to 3170; -	402	α-*proteobacteria Roseobacter*	54%/ 70%	periplasmic peptide-binding protein
2-dehydropantoate 2-reductase	5376 to 6299; +	307	B-*proteobacteria: Polaromonas Actinobacteria:*	36%/ 55%	Required for synthesis of thiamine via APB biosynthetic pathway
conserved hypothetical protein	7590 to 6364; -	408	α-*proteobacteria: Roseobacter*	59%/ 70%	Carbohydrate transport and metabolism
amidohydrolase family protein	7930 to 9069; +	379	α-*proteobacteria: Roseobacter*	70%/ 81%	Metal-dependent amidase/ aminoacylase/ carboxypeptidase
TPR repeat	11078 to 10170; -	302	*Bacteroidetes: Psychroflexus*	30%/ 56%	Function unknown
oxidoreductase, GMC family protein	**11979 to 13319; +**		α-*proteobacteria: Roseobacter*		Amino Acid transport and metabolism (specific function unknown)
Mandelate racemase/ muconate lactonizing enzyme	13330 to 14430; +	366	α-*proteobacteria: Roseobacter*	76%/ 89%	involved in aromatic acid catabolism (specific function unknown)
TPR domain protein	14467 10 16119; +	550	α-*proteobacteria: Roseobacter*	64%/ 78%	Function unknown
hypothetical protein ROS217_06184	17187 to 16885; -	100	γ-*proteobacteria: Vibrio vulnificus*	44%/ 56%	Function unknown
hypothetical protein SKA53_07981	17366 to 17821; +	151	*Bacteroidetes: Psychroflexus*	45%/ 64%	Function unknown
integrative genetic element Gsu5, resolvase	18253 to 17831; -	140	α-*proteobacteria: Roseobacter*	38%/ 57%	
Cyclic nucleotide-binding domain (cNMP-BD) protein	19310 to 18969; -	113	*a-proteobacteria: Rhodospirillum* γ-*proteobacteria:*	38%/ 56%	gene activator and regulatory subunit
Metal dependant beta lactamase protein	19997 to 19311; -	228	δ-*proteobacteria: Cyanobacteria:* α-*proteobacteria:*	30%/ 50%	
Adenylyl cyclase: **CYCc**	**20327 to 21616; +**	429	γ-*proteobacteria: Alcanivorax*	42%/64%	Catalyses the formation of cyclic AMP (cAMP)
5S Ribosomal subunit					
23S Ribosomal subunit					
ITS region					
16S Ribosomal subunit					
Ferric reductase-like transmembrane component-like	28420 to 27971; -	149	α-*proteobacteria: Roseobacter*	57%/ 74%	may be capable of moving electrons across the plasma membrane
Oxidoreductase molybdopterin binding domain	29404 to 28538; -	228	α-*proteobacteria: Roseobacter*	74%/ 84%	catalyses oxidation of sulphite to sulphate
AAA ATPase	**32219 to 29540; -**		α-*proteobacteria: Roseobacter*		Assists in the assembly, operation, or disassembly of protein complexes

Table 2: Protein coding sequences on PBAC 250.

Gene name	PBAC 248	*Roseovarious nubinhibens*	*Silicibacter pomeryi*	*Silicibacter TM1040*	*Sulfitobacter sp. NAS 14.1*	*Sulfitobacter sp. EE-46*	*Jannascria sp. CCS1*	*Rhoderbacter sphaeroides*
ABC_tran	51.82%	62.00%	63.00%	62.00%	60.00%	60.00%	62.00%	70.00%
BPD_transp_1	50.38%	61.00%	60.00%	61.00%	59.00%	59.00%	62.00%	66.00%
BPD_transp_1	52.50%	60.00%	60.00%	58.00%	59.00%	59.00%	62.00%	66.00%
SBP_bac_5	49.55%	62.00%	63.00%	62.00%	60.00%	60.00%	63.00%	69.00%
AGE	52.89%	-	-	-		-	-	70.00%
Peptidase_M20	52.37%	-	69.00%	-		-	62.00%	-
BetA	52.32%	-	-	-		-	-	-
MR_like	51.14%	67.00%	66.00%	64.00%		-	-	-
TPR	51.66%	65.00%	66.00%	53.00%		-	-	-
Hyp. (ferric)	51.56%	67.00%	63.00%	57.00%	60.00%	58.00%	61.00%	67.00%
Oxidored_molyb	50.52%	62.00%	61.00%	58.00%	58.00%	58.00%	62.00%	65.00%
AAA	54.40%		63.00%	61.00%	60.00%	61.00%	63.00%	67.00%
Genome G-C estimate	51.50%	63.30%	63.60%	59.60%	59.43%	59.30%	62.10%	67.50%
Actual genome G-C content	-	63.91%	64.07%	59.87%	60.05%	60.31%	62.24%	68.24%
Difference	-	-0.61%	-0.47%	-0.27%	-0.62%	-1.01%	-0.14%	-0.74%

Gene name	PBAC 250	*Roseovarious nubinhibens*	*Silicibacter Pomeryi*	*Silicibacter TM1040*	*Sulfitobacter sp. NAS 14.1*	*Sulfitobacter sp. EE-46*	*Jannascria sp. CCS1*	*Rhoderbacter sphaeroides*
C-5 epimerase	49.73%	67.00%	-	-		-	-	-
Sulfotransfer_1	52.87%	-	-	-		-	59.00%	-
WcaJ	51.02%	-	-	-		62.00%	-	67.00%
Hyp. (Wzz)	52.02%	63.00%	-	-	56.00%	57.00%	-	69.00%
Glycos_transf_N	51.82%	67.00%	67.00%	65.00%	65.00%	65.00%	64.00%	74.00%
LpxK	54.97%	72.00%	68.00%	68.00%	65.00%	64.00%	68.00%	75.00%
Gcv_T	50.26%	64.00%	61.00%	59.00%	62.00%	61.00%	-	-
Genome G-C estimate	51.8%	66.60%	66.00%	64.00%	62.00%	62.20%	63.67%	71.25%
Actual genome G-C content	-	63.91%	64.07%	59.87%	60.05%	60.31%	62.24%	68.24%
Difference	-	+2.69%	+1.93%	+4.13%	+1.95%	+1.49%	+1.43%	+3.01%