# QUANTITATIVE ANALYSIS OF BACTERIAL DMSP-DEGRADING GENE DIVERSITY, ABUNDANCE, AND EXPRESSION IN MARINE SURFACE WATER ENVIRONMENTS

by

#### VANESSA ANN VARALJAY

#### (Under the Direction of Mary Ann Moran)

#### ABSTRACT

Dimethylsulfoniopropionate (DMSP) is an algal-derived sulfur compound degraded by marine bacterioplankton in surface waters via two competing mechanisms: (1) DMSP cleavage that produces the climatically-relevant sulfur gas dimethylsulfide (DMS); and (2) DMSP demethylation that produces less-volatile intermediates used for amino acid synthesis or energy. Given the implications for climate, the discovery of the genes mediating the first steps in each pathway, *dddP* (DMSP cleavage) and *dmdA* (DMSP demethylation), provided the basis to quantify the diversity, abundance, and expression of this genetic capability in the environment. Using an *in silico* pipeline based on marine metagenomic data, 10 primer sets for environmental clades of *dmdA* were iteratively optimized to target sequences in a southeastern U.S. coastal ocean. Using pyrosequencing, >700 total sequence clusters (~90% sequence similarity) were retrieved and thousands of sequences per primer exhibited specificity to the correct group, signifying the importance of environmental sequence data in primer design. The vetted primer sets were used in quantitative PCR in the North Pacific Ocean over 10 months in the upper and lower euphotic zones. DMSP-degrading genes were more abundant at the surface (maximum of  $\sim 16\%$  (*dmdA*) and  $\sim 2\%$  (*dddP*) of cells harboring a gene), consistent with higher DMSP concentrations, temperature, and solar radiation. The *dmdA* pool was dominated by genes from the SAR11 clade with lesser but consistent contributions of roseobacters and Gammaproteobacteria. In Monterey Bay, an autonomous sampling instrument captured near-daily abundance and expression of DMSP-degrading genes from two taxa, roseobacters (*dmdA* and *dddP*) and SAR11 (*dmdA*) during highly variable DMSP concentrations due to sequential phytoplankton blooms. SAR11 genes were more abundant than Roseobacter genes, but expression levels per gene were lower, particularly in particle-associated vs. free-living fractions. Expression ratios for all genes were best correlated with particulate DMSP relative to the total phytoplankton pool, a proxy for high-DMSP-producing phytoplankton. Overall, PCR analysis with environmentally-relevant primer sets indicated DMSP-degrading genes are abundant in marine surface waters. Taxon-specific spatial and temporal trends in gene diversity, abundance, and expression in association with DMSP, phytoplankton, and other environmental variables were observed, providing new insights into the bacteria that cycle sulfur in the surface ocean.

INDEX WORDS: Sulfur, DMSP, DMSP demethylation, Marine bacteria, DmdA,
 DddP, Quantitative PCR, Pyrosequencing, Metagenomics, Primer
 design, Oligotrophic, North Pacific, Autonomous remote sensing,
 Coastal mooring, Monterey Bay, Phytoplankton, SAR11,
 Roseobacters, Gene expression, Particle-associated, Free-living

# QUANTITATIVE ANALYSIS OF BACTERIAL DMSP-DEGRADING GENE DIVERSITY, ABUNDANCE, AND EXPRESSION IN MARINE SURFACE WATER ENVIRONMENTS

by

### VANESSA ANN VARALJAY

B.S., Cleveland State University, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

© 2012

Vanessa Ann Varaljay

All Rights Reserved

# QUANTITATIVE ANALYSIS OF BACTERIAL DMSP-DEGRADING GENE DIVERSITY, ABUNDANCE, AND EXPRESSION IN MARINE SURFACE WATER ENVIRONMENTS

by

## VANESSA ANN VARALJAY

Major Professor: Mary Ann Moran

Committee:

Brian Binder Anna Karls William Whitman

Electronic Version Approved:

Maureen Grasso Dean of the Graduate School The University of Georgia May 2012

# DEDICATION

To my family who always believed in me.

#### ACKNOWLEDGEMENTS

First and foremost, I thank my advisor, Mary Ann Moran, whose guidance and support has been essential to my graduate school experience and success. I could not have asked for a better mentor. I will always be grateful to Mary Ann for teaching me to be an independent researcher, helping me to become a better writer, and preparing me for a future in science. I also thank my committee members, Brian Binder, Anna Karls and Barny Whitman, who not only challenged me but provided me with helpful advice and suggestions for improving and developing my research.

There are a number of people that I have collaborated with from various institutions over the past six years that I am thankful for having had the chance to work with and learn from: Naomi Levine, Sam Wilson, David Karl, Christina Preston, Julie Robidart, Roman Marin, Chris Scholin, and Ron Kiene.

In addition, I thank all of the lab members (former and present), colleagues, staff, and friends at UGA that have supported me during my graduate career, with special regards to: Scott Gifford, Erinn Howard, Leanne Powers, Carrie Givens, Bryndan Durham, Christine Hladik, Maria Vila, Ryan Newton, Johanna Rinta-Kanto, Chris Reisch, Bradley Tolar, Brandon Satinsky, Noah Lawler, Joanna Green, Heather Reader, Marcia Hsu, Shalabh Sharma, Erin Biers, Shulei Sun, and Wendy Ye. Many thanks goes to Camille English, who made getting things done in lab very efficient; Sharon Barnhart, who was essential for assisting with all things technical; and the entire office staffs in the Marine Science and Microbiology departments for their help.

V

I also acknowledge two exceptional individuals, the members of my incoming 2006 class, Alecia Septer and Sue Hennigan. I am lucky I have had the opportunity to get to know them both personally and professionally.

Finally, a big thank you and appreciation goes to my amazing family: my mom, Virginia, my dad, Nicholas, and my brother, Max. I owe much of my success to their unconditional love and support. I also thank my fiancé, Matthew Ray, who was always there to listen to me, inspire me, and make me smile every day.

## TABLE OF CONTENTS

	Page
ACKNOV	VLEDGEMENTSv
СНАРТЕ	R
1	INTRODUCTION AND LITERATURE REVIEW1
2	DEEP SEQUENCING OF A DIMETHYLSULFONIOPROPIONATE
	DEGRADING GENE (DMDA) BY USING PCR PRIMERS DESIGNED ON
	THE BASIS OF MARINE METAGENOMIC DATA25
3	BACTERIAL DIMETHYLSULFONIOPROPIONATE DEGRADATION
	GENES IN THE OLIGOTROPHIC NORTH PACIFIC SUBTROPICAL
	GYRE61
4	IN SITU TIME-SERIES MEASUREMENTS OF BACTERIAL
	DIMETHYLSULFONIOPROPIONATE GENES FROM A MONTEREY
	BAY COASTAL MOORING103
5	CONCLUSIONS158

## APPENDICES

А	CHAPTER 2 SUPPLEMENTARY MATERIAL	163
В	CHAPTER 3 SUPPLEMENTARY MATERIAL	172
С	CHAPTER 4 SUPPLEMENTARY MATERIAL	183

#### CHAPTER 1

#### INTRODUCTION AND LITERATURE REVIEW

#### I. DMSP significance

Dimethylsulfoniopropionate (DMSP) is a reduced organic sulfur compound produced by marine phytoplankton and some marine plant species for use primarily as an osmolyte (12, 48, 50, 81). DMSP was first isolated in the marine alga *Polysiphonia* sp. by Challenger and Simpson in 1948 (11) and has since been shown to also function as a predator deterrent, antioxidant, chemoattractant, and cyroprotectant (28, 47, 64, 71, 87). Some phytoplankton species cleave DMSP using DMSP lyases (69, 88) to produce dimethylsulfide (DMS) which can act as an antioxidant and signaling molecule (46, 71). The DMS produced is also the most abundant organic sulfur gas emitted to the atmosphere globally (accounting for 28 Tg S<sup>-1</sup> y<sup>-1</sup>; ref. 44) and, once oxidized to sulfur dioxide and sulfate aerosols, becomes the precursor to cloud condensation nuclei (1-3, 5, 6, 39, 40). The role of DMS as a global climate regulator was presented by Charlson et al. (13) in 1987 as the "CLAW" hypothesis (CLAW stands for the first letter of each of the authors' last names) which suggested that phytoplankton, DMS, and solar radiation are entwined in a negative feedback loop.

Further research has provided insight in support of or against the climactic impacts of light, DMS, and cloud albedo (39, 54, 78, 82), but predicting global DMS flux has proven largely unsuccessful when based on bulk measurements of chlorophyll and DMS concentrations (29, 65). Early on, phytoplankton were identified as the primary

sources of DMS production from DMSP (10, 26), and while phytoplankton are indeed significant DMS producers, the systematic production and release of DMS to the atmosphere is contingent on a number of complicating factors (70). For instance, physical processes such as water column mixing and photolysis can affect DMS yield and degassing (67, 68, 77). Additionally, both DMSP and DMS production are constrained to certain phytoplankton taxa, such as dinoflagellates, prymnesiophytes, and diatoms, which vary in abundance, distributions, and DMSP production rate (8, 62, 80). Finally, it was shown that biological degradation by marine bacterioplankton is responsible for a significant amount of DMSP and DMS cycling in surface waters (37, 38, 41, 56, 72).

The majority of DMSP produced by phytoplankton is released into the seawater as either dissolved DMS or DMSP by cell death, zooplankton grazing (18), and viral lysis (22) where it is available for marine bacterial consumption (86, 91). The dissolved DMS pool can be used for carbon or sulfur source or degraded to DMSO by specialized bacterial groups (63, 84). The dissolved DMSP pool is rapidly consumed (rates on the order of hours to days; ref. 32) and may initially be accumulated intracellularly by marine bacteria to use as an osmolyte (19, 58). Subsequently, the DMSP is degraded by one of two enzymatically mediated and competing pathways (89) for which some bacteria are able to carry out one pathway, some the other, and some both (4, 20, 36, 43). Bacteria that harbor the DMSP cleavage pathway produce DMS, most of which diffuses into the seawater (30). With the DMSP demethylation pathway, bacteria produce methanethiol (MeSH) which can either be used to produce methionine or cysteine for protein synthesis or be further oxidized (34, 59). The demethylation pathway is a bacterial-specific process which, depending on environment and season, directs up to 50-90% of the

DMSP-derived sulfur away from DMS production and atmospheric S flux (31, 33, 34, 66, 91). Therefore, bacterial processing of DMSP is responsible for routing a large portion of sulfur into the microbial food web rather than to the atmosphere.

In both demethylation and cleavage pathways, a DMSP-derived 3-carbon compound is provided for the bacteria, with the critical difference being whether the methyl and sulfur groups of the molecule are used or lost (33). Several studies have suggested that after reaching a certain sulfur quota, degradation switches from demethylation to cleavage, which still provides carbon to the organism (20, 33, 34). Indeed, DMSP can satisfy ~100% of bacterial sulfur demands (34). However, regulation and environmental factors in the bacterial switch between demethylation and cleavage have not been clearly demonstrated. Given the importance of bacteria in DMSP degradation and the need to better understand the regulation of these two degradation pathways, recent advances were made in identifying the genes involved in each.

#### **II. Bacterial DMSP degrading genes**

The first gene in the demethylation pathway, the DMSP demethylase gene (*dmdA*), was discovered by Howard et al. (23) in 2006 using transposon mutagenesis in the cultured marine roseobacter, *Ruegeria* (formerly *Silicibacter*) *pomeroyi* DSS-3. The *dmdA* gene encodes for the first step in the demethylation pathway (Fig. 1.1), where 3-methiolpropionate (MMPA) is produced when one of the methyl groups of DMSP is removed and transferred to a tetrahydrofolate (THF) carrier (23, 58). This first step is critical in eliminating the possibility that DMSP will be degraded instead to DMS. The *dmdA* gene was identified in the genomes of other marine roseobacters and in the

genomes of the ubiquitous SAR11 (45) *Candidatus* 'Pelagibacter ubique' strains, for which the gene was also experimentally verified (23).

At this same time, large-scale marine metagenomic sequence data were becoming increasingly available and expanded the knowledge of *dmdA* abundance and diversity. The GOS metagenomes released in 2005 (83) and 2007 (61) comprised more than 7.7 million reads from DNA samples collected from the surface waters of the Sargasso Sea, the northwest Atlantic, the Eastern U.S. coast, and the Pacific Ocean. From the 2007 GOS metagenome (61), a second study by Howard et al. (24) in 2008 found 1,701 *dmdA* sequences and estimated *dmdA* to be harbored by half of surface-water bacteria. Since then, recent surveys of metagenomic data have confirmed the high *dmdA* abundance in marine bacterial communities (44, 55).

Sequences from marine metagenomic data and cultured marine bacteria group into 5 protein clades: A, B, C, D, and E (24). The Roseobacter and Rhodospirillales *dmdA* genes belong to Clade A, SAR11 *dmdA* genes to Clades C and D, and the Gammaproteobacteria HTCC2080 *dmdA* gene to Clade E. The *dmdA* homolog identified from the sequenced genome of a SAR116 organism, *Candidatus* 'Puniceispirillum marinum' IMCC1322, belongs to Clade B, which until recently (25) did not have a cultured representative. A small proportion of the sequences falls outside of these clades and are thus unclassified (24). Although fairly cohesive at the protein level with at least one cultured representative per clade, the clades appear to represent heterogeneous subgroups of marine bacteria possessing *dmdA* genes. These findings are indicative of the diverse marine bacterial taxa (Roseobacter, SAR11, Gammaproteobacteria, SAR116,

and uncultured groups) that have DMSP demethylases and signify an important ecological role for *dmdA* in natural communities.

Given the importance and widespread abundance of DMSP demethylation, the genes encoding the subsequent steps of the demethylation pathway were identified and characterized in *R. pomeroyi* DSS-3 by Reisch et al. (59) in 2011. The genes, designated *dmdB*, -*C*, and –*D*, mediate the degradation of MMPA to acetaldehyde in a novel fatty acid oxidation-like pathway. In the first step, MMPA-CoA thioester is produced from MMPA by DmdB (Fig. 1.1). MMPA-CoA is then oxidized by DmdC, forming a double bond and producing methylthiolacrylyl-CoA (MTA-CoA) (Fig. 1.1). Lastly, MTA-CoA is hydrated by DmdD to produce acetaldehyde, CO<sub>2</sub>, and free CoA (Fig. 1.1). The *dmdB* and *dmdC* genes were found in abundance in the genomes of roseobacters and SAR11 bacteria, marine metagenomic data, and even bacteria not harboring *dmdA* genes (such as *Burkholderia* sp.). However, homologs to the *dmdD* gene were not abundant and a functional homolog was not identified in SAR11, suggesting that an important isofunctional gene still remains to be identified.

In contrast to the first step of DMSP demethylation, which is mediated by one gene, analysis of the first step of DMSP cleavage has yielded 6 different genes (*dddD*, *dddL*, *dddP*, *dddQ*, *dddY*, and *dddW*; Fig. 1.1). All of the genes mediate the production of DMS, although by widely different enzyme types (16). Gene identification in the DMSP cleavage pathway was first investigated in the marine Gammaproteobacterium *Marinomonas* sp. by screening cosmid fragments of its genome and expressing them in *E.coli* when grown on DMSP and carbon (76). This led to the discovery of the *dddD* gene, which encodes for an acyl-coA transferase that likely produces a CoA intermediate

and yields beta-hydroxypropionate and DMS (57, 76) (Fig. 1.1). The *dddD* gene was found in some roseobacter genomes, but other cultured members known to produce DMS did not have a homolog. This led to the investigation of other DMSP cleavage genes. The second discovered gene, *dddL*, a cupin domain enzyme, was not homologous to *dddD* or any polypeptide known (14). It was deemed a "true" DMSP lyase as it conferred the ability to produce acrylate and DMS from DMSP (14) (Fig. 1.1). The third gene, *dddP*, initially characterized as a metallopeptidase although it does not function as such, produces DMS and acrylate (35, 73) (Fig. 1.1). *dddP* has been found in roseobacters, SAR116, and even some fungi. The fourth gene, *dddQ*, a second cupin domain DMSP lyase, was also identified in roseobacters (including those also harboring a *dddP*) and uncultured bacteria (74) (Fig. 1.1). The *dddW* gene was identified as the third cupin domain DMSP lyase also found in roseobacters (75) (Fig. 1.1). However, *dddL*, *dddQ*, and *dddW* share no sequence similarity to one another (75). Lastly, the *dddY* gene is a periplasmic DMSP lyase found in Beta- and Gammaproteobacteria where it may be important in DMS production in marine sediments (15) (Fig. 1.1).

The *ddd* genes are rarer than *dmdA* in the metagenomic data, as all *ddd* genes together are present in less than 10% of bacterial cells. In a recent survey of the GOS metagenome, *dddY* was completely absent, *dddD*, *-L*, *-Q*, *-W* were present in less than 1% of cells. *dddP*, the most abundant gene thus far, was present in ~6% of cells (44). The paucity of these genes in the environment could be attributed to a number of factors. Perhaps the most abundant DMS-producing gene has not yet been found or marine bacteria are simply not major players in this process. While DMSP cleavage genes are

present in marine roseobacters, even multiple *ddd* types in the same genome, they are absent in the abundant SAR11 bacteria.

Knowing the genes and pathways harbored by various taxa has important implications in understanding how DMSP is processed. Tripp et al. (79) demonstrated that SAR11 requires reduced sulfur for growth, assimilating DMSP sulfur, presumably through the DMSP demethylation pathway without the capability of DMSP cleavage. On the other hand, roseobacters do not require reduced sulfur for growth and are capable of both DMSP degradation pathways (20, 42). In addition, Reisch et al. (58) found that cultured Roseobacter and SAR11 DmdAs are somewhat different in DMSP turnover rates and Km values. In the environment, these differences in pathway capabilities and enzyme properties may be relevant during seasonal phytoplankton successions or blooms in various water types, for example in coastal vs. open ocean systems where bacterial taxa, DMSP, DMS, carbon, nutrients, and other environmental parameters differ (9, 17, 21, 51, 52, 80, 90). Differing ecological DMSP degradation strategies may be used in such cases by members of the bacterioplankton community harboring demethylation and/or cleavage genes.

Culture-based gene discovery, microarrays, and metagenomic and metatranscriptomics studies (7, 60, 85) have provided the genetic foundation to investigate bacterial regulation of DMSP catabolism. The discoveries of key genes encoding the first steps of competing DMSP-degrading pathways for demethylation (*dmdA*) and DMS production (*dddP*; the most abundant DMS producing gene), along with available sequence and biogeochemical data, now allow for the development of

molecular tools to quantitatively study their distributions, diversity, abundance, and expression in field studies.

#### **III.** Objectives

In Chapter 2, metagenomic-based primer design and deep sequencing was used to explore the DMSP demethylase gene, *dmdA*, from a southeastern coastal bacterial community. This study used thousands of marine metagenomic sequences to rigorously design and optimize quantitative PCR primer sets *in silico* and apply the primers to examine diversity from environmental Roseobacter, SAR11, and Gammaproteobacteria *dmdA* subclades. Thousands of amplicon sequences were used to assess the specificity and applicability of the metagenomic-based primer sets in the field and to compare *dmdA* sequence diversity from free-living and particle-associated size fractions.

In Chapter 3, the suite of vetted and metagenomic-based primer sets were then used in a 10-month time-series in the North Pacific Subtropical Gyre, one of the largest of the world's biomes (27), to quantitatively track the abundance and expression of *dmdA* and *dddP* genes. Gene distributions were followed temporally at two ecologically distinct depths, at 25 m in the nutrient-limited upper euphotic zone and at ~125 m corresponding to the deep chlorophyll maximum (DCM) in the nutrient-replete lower euphotic zone. In this study, biochemical and molecular sampling were measured concurrently to ascertain the physical and environmental parameters which covary with bacterial DMSP-degrading genes and to understand how this shapes DMSP cycling in the system.

In Chapter 4, *in situ* autonomous real-time monitoring (53) was used to resolve Roseobacter and SAR11 DMSP-degrading gene abundance and expression over near-

daily scales coinciding with measurements of phytoplankton, chlorophyll, and DMSP variability in Monterey Bay, a dynamic coastal upwelling system (49). The use of internal extraction standards and traditional sampling methods (Niskin bottle) were used to assess and validate measurements from remotely collected samples. Quantifying the expression of *dmdA* and *dddP* genes with homology to Roseobacter clade strain HTCC2255 provided the opportunity to study competing DMSP-degrading pathways during variable DMSP fluxes. In addition, differences in abundance and gene expression ratios were compared between the Roseobacters and SAR11 genes in whole water and in particle vs. free-living size-fractions to assess niche differentiation in these co-occurring DMSP-degrading bacteria.

#### References

- Andreae, M. O. 1990. Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. Mar. Chem. 30:1-29.
- Andreae, M. O., and P. J. Crutzen. 1997. Atmospheric aerosols: biogeochemical sources and role in atmospheric chemistry. Science 276:1052-1058.
- 3. Andreae, M. O., and H. Raemdonck. 1983. Dimethyl sulfide in the surface ocean and the marine atmosphere: a global view. Science **221**:744-747.
- Ansede, J. H., R. Friedman, and D. C. Yoch. 2001. Phylogenetic analysis of culturable dimethyl sulfide-producing bacteria from a *Spartina*-dominated salt marsh and estuarine water. Appl. Environ. Microbiol. 67:1210-1217.
- 5. Bates, T. S., R. J. Charlson, and R. H. Gammon. 1987. Evidence for the climatic role of marine biogenic sulphur. Nature **329**:319-321.
- Berresheim, H., F. L. Eisele, D. J. Tanner, L. M. McInnes, D. C. Ramseybell, and D. S. Covert. 1993. Atmospheric sulfur chemistry and cloud condensation nuclei (CCN) concentrations over the northeastern pacific coast. J. Geophys. Res-Atmos. 98:12701-12711.
- Bürgmann, H., E. C. Howard, W. Ye, F. Sun, S. Sun, S. Napierala, and M. A. Moran. 2007. Transcriptional response of *Silicibacter pomeroyi* DSS-3 to dimethylsulfoniopropionate (DMSP). Environ. Microbiol. 9:2742-2755.
- Burkill, P. H., S. D. Archer, C. Robinson, P. D. Nightingale, S. B. Groom, G.
   A. Tarran, and M. V. Zubkov. 2002. Dimethyl sulphide biogeochemistry within

a coccolithophore bloom (DISCO): an overview. Deep-Sea Res. Pt II. **49**:2863-2885.

- Cantin, G., M. Levasseur, S. Schultes, and S. Michaud. 1999. Dimethylsulfide (DMS) production by size-fractionated particles in the Labrador Sea. Aquat. Microb. Ecol. 19:307-312.
- Cantoni, G., and D. Anderson. 1956. Enzymatic cleavage of dimethylpropiothetin by *Polysiphonia lanosa*. J. Biol. Chem. 222:171-177.
- 11. Challenger, F., and M. Simpson. 1948. Studies on biological methylation. Part
  12. A precursor of the dimethyl sulphide evolved by *Polysiphonia fastigiata*.
  Dimethyl-2-carboxyethylsulphonium hydroxide and its salts J. Chem. Soc.
  3:1591-1597.
- Chambers, S. T., C. M. Kunin, D. Miller, and A. Hamada. 1987.
   Dimethylthetin can substitute for glycine betaine as an osmoprotectant molecule for *Escherichia coli*. J. Bacteriol. 169:4845-4847.
- Charlson, R. J., J. E. Lovelock, M. O. Andreae, and S. G. Warren. 1987.
   Oceanic phytoplankton, atmospheric sulfur, cloud albedo, and climate. Nature 326:655-661.
- 14. Curson, A. R. J., R. Rogers, J. D. Todd, C. A. Brearley, and A. W. B.
   Johnston. 2008. Molecular genetic analysis of a dimethylsulfoniopropionate
   lyase that liberates the climate-changing gas dimethylsulfide in several marine α proteobacteria and *Rhodobacter sphaeroides*. Environ. Microbiol. 10:757-767.

- Curson, A. R. J., M. J. Sullivan, J. D. Todd, and A. W. B. Johnston. 2011.
   DddY, a periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse species of *Proteobacteria*. ISME J. 5:1191-1200.
- Curson, A. R. J., J. D. Todd, M. J. Sullivan, and A. W. B. Johnston. 2011. Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. Nat Rev Micro 9:849-859.
- Dacey, J. W. H., F. A. Howse, A. F. Michaels, and S. G. Wakeham. 1998. Temporal variability of dimethylsulfide and dimethylsulfoniopropionate in the Sargasso Sea. Deep-Sea Res. Pt I. 45:2085-2104.
- Dacey, J. W. H., and S. G. Wakeham. 1986. Oceanic dimethylsulfide: production during zooplankton grazing on phytoplankton. Science 233:1314-1316.
- Diaz, M. R., P. T. Visscher, and B. F. Taylor. 1992. Metabolism of dimethylsulfoniopropionate and glycine betaine by a marine bacterium. FEMS Microbiol. Lett. 96:61-65.
- González, J. M., R. P. Kiene, and M. A. Moran. 1999. Transformation of sulfur compounds by an abundant lineage of marine bacteria in the α-subclass of the class *Proteobacteria*. Appl. Environ. Microbiol. 65:3810-3819.
- González, 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. Appl. Environ. Microbiol. 66:4237-4246.

- Hill, R. W., B. A. White, M. T. Cottrell, and J. W. H. Dacey. 1998. Virusmediated total release of dimethylsulfoniopropionate from marine phytoplankton: a potential climate process. Aquat. Microb. Ecol. 14:1-6.
- Howard, E. C., J. R. Henriksen, A. Buchan, C. R. Reisch, H. Bürgmann, R. Welsh, W. Y. Ye, J. M. González, K. Mace, S. B. Joye, R. P. Kiene, W. B. Whitman, and M. A. Moran. 2006. Bacterial taxa that limit sulfur flux from the ocean. Science 314:649-652.
- Howard, E. C., S. Sun, E. J. Biers, and M. A. Moran. 2008. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. Environ. Microbiol. 10:2397-2410.
- 25. Howard, E. C., S. Sun, C. R. Reisch, D. A. del Valle, H. Bürgmann, R. P. Kiene, and M. A. Moran. 2011. Changes in dimethylsulfoniopropionate demethylase gene assemblages in response to an induced phytoplankton bloom. Appl. Environ. Microbiol. 77:524-531.
- Iverson, R. L., F. L. Nearhoof, and M. O. Andreae. 1989. Production of dimethylsulfonium propionate and dimethylsulfide by phytoplankton in estuarine and coastal waters. Limnol. Oceanogr. 34:53-67.
- Karl, D. M. 1999. A sea of change: biogeochemical variability in the North Pacific Subtropical Gyre. Ecosystems 2:181-214.
- 28. Karsten, U., K. Kuck, C. Vogt, and G. O. Kirst. 1996.
   Dimethylsulfoniopropionate production in phototrophic organisms and its physiological function as a cryoprotectant, p 143-153. *In* R.P. Kiene, P.T.

Visscher, M.D. Keller, and G.O. Kirst (eds), Biological and environmental chemistry of DMSP and related sulfonium compounds. New York: Plenum Press.

- 29. Kettle, A. J., M. O. Andreae, D. Amouroux, T. W. Andreae, T. S. Bates, et al. 1999. A global database of sea surface dimethylsulfide (DMS) measurements and a procedure to predict sea surface DMS as a function of latitude, longitude, and month. Global Biogeochem. Cy. 13:399-444.
- Kiene, R. P. 1990. Dimethyl sulfide production from dimethylsulfoniopropionate in coastal seawater samples and bacterial cultures. Appl. Environ. Microbiol. 56:3292-3297.
- Kiene, R. P. 1996. Production of methanethiol from dimethylsulfoniopropionate in marine surface waters. Mar. Chem. 54:69-83.
- 32. Kiene, R. P., and L. J. Linn. 2000. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico Limnol. Oceanogr. 45:849-861.
- Kiene, R. P., L. J. Linn, and J. A. Bruton. 2000. New and important roles for DMSP in marine microbial communities. J. Sea Res. 43:209-224.
- 34. Kiene, R. P., L. J. Linn, J. Gonzalez, M. A. Moran, and J. A. Bruton. 1999. Dimethylsulfoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. Appl. Environ. Microbiol. 65:4549-4558.
- 35. **Kirkwood, M., N. E. Le Brun, J. D. Todd, and A. W. B. Johnston**. The *dddP* gene of *Roseovarius nubinhibens* encodes a novel lyase that cleaves

dimethylsulfoniopropionate into acrylate plus dimethyl sulfide. Microbiology **156**:1900-1906.

- Ledyard, K. M., and J. W. H. Dacey. 1994. Dimethylsulfide production from dimethylsulfoniopropionate by a marine bacterium. Mar. Ecol-Prog. Ser. 110:95-103.
- Ledyard, K. M., and J. W. H. Dacey. 1996. Microbial cycling of DMSP and DMS in coastal and oligotrophic seawater. Limnol. Oceanogr. 41:33-40.
- Ledyard, K. M., E. F. DeLong, and J. W. H. Dacey. 1993. Characterization of a DMSP-degrading bacterial isolate from the Sargasso Sea. Arch. Microbiol. 160:312-318.
- Lovelock, J., R. Maggs, and R. Rasmussen. 1972. Atmospheric dimethyl sulfide and the natural sulfur cycle. Nature 237:452-453.
- 40. **Malin, G., and G. O. Kirst.** 1997. Algal production of dimethyl sulfide and its atmospheric role. J. Phycol. **33**:889-896.
- Malmstrom, R. R., R. P. Kiene, M. T. Cottrell, and D. L. Kirchman. 2004. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic Ocean. Appl. Environ. Microbiol. 70:4129-4135.
- 42. Moran, M. A., R. Belas, M. A. Schell, J. M. Gonzalez, F. Sun, et al. 2007.
  Ecological genomics of marine roseobacters. Appl. Environ. Microbiol. 73: 455969.

- 43. Moran, M. A., J. M. Gonzalez, and R. P. Kiene. 2003. Linking a bacterial taxon to sulfur cycling in the sea: Studies of the marine Roseobacter group.
  Geomicrobiol. J. 20:375-388.
- 44. **Moran, M. A., C. R. Reisch, R. P. Kiene, and W. B. Whitman**. 2012. Genomic insights into bacterial DMSP transformations. Ann. Rev. Mar. Sci. **4**:523-524.
- 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.
- 46. Nevitt, G. A., R. R. Veit, and P. Kareiva. 1995. Dimethyl sulphide as a foraging cue for Antarctic *Procellariiform* seabirds. Nature **376**:680-682.
- 47. Nishiguchi, M. K., and G. N. Somero. 1992. Temperature- and concentrationdependence of compatibility of the organic osmolyte β-dimethylsulfoniopropionate. Cryobiology 29:118-124.
- 48. Otte, M. L., G. Wilson, J. T. Morris, and B. M. Moran. 2004.
  Dimethylsulphoniopropionate (DMSP) and related compounds in higher plants. J.
  Exp. Bot. 55:1919-1925.
- Pennington, J. T., and F. P. Chavez. 2000. Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989-1996 in Monterey Bay, California. Deep-Sea Res. Pt II. 47:947-973.
- 50. Pichereau, V., J.-A. Pocard, J. Hamelin, C. Blanco, and T. Bernard. 1998. Differential effects of dimethylsulfoniopropionate, dimethylsulfonioacetate, and other s-methylated compounds on the growth of *Sinorhizobium meliloti* at low and high osmolarities. Appl. Environ. Microbiol. 64:1420-1429.

- Pinhassi, J., M. M. Sala, H. Havskum, F. Peters, O. Guadayol, A. Malits, and C. Marrase. 2004. Changes in bacterioplankton composition under different phytoplankton regimens. Appl. Environ. Microbiol. 70:6753-6766.
- Pinhassi, J., R. Simó, J. M. González, M. Vila, L. Alonso-Sáez, R. P. Kiene, M. A. Moran, and C. Pedrós-Alió. 2005. Dimethylsulfoniopropionate turnover is linked to the composition and dynamics of the bacterioplankton assemblage during a microcosm phytoplankton bloom. Appl. Environ. Microbiol. 71:7650-7660.
- 53. Preston, C. M., A. Harris, J. P. Ryan, B. Roman, R. Marin, III, S. Jensen, C. Everlove, J. Birch, J. M. Dzenitis, D. Pargett, M. Adachi, K. Turk, J. P. Zehr, and C. A. Scholin. 2011. Underwater application of quantitative pcr on an ocean mooring. PLoS ONE 6:e22522.
- 54. **Quinn, P. K., and T. S. Bates**. 2011. The case against climate regulation via oceanic phytoplankton sulphur emissions. Nature **480**:51-56.
- 55. Raina, J.-B., E. A. Dinsdale, B. L. Willis, and D. G. Bourne. 2010. Do the organic sulfur compounds DMSP and DMS drive coral microbial associations? Trends Microbiol. 18:101-108.
- Raina, J.-B., D. Tapiolas, B. L. Willis, and D. G. Bourne. 2009. Coralassociated bacteria and their role in the biogeochemical cycling of sulfur. Appl. Environ. Microbiol. 75:3492-3501.
- Reisch, C. R., M. A. Moran, and W. B. Whitman. 2011. Bacterial catabolism of dimethylsulfoniopropionate. Front. Microbiol. 2:172.

# Reisch, C. R., M. A. Moran, and W. B. Whitman. 2008. Dimethylsulfoniopropionate-dependent demethylase (DmdA) from *Pelagibacter* ubique and *Silicibacter pomerovi*. J. Bacteriol. 190:8018-8024.

- 59. Reisch, C. R., M. J. Stoudemayer, V. A. Varaljay, I. J. Amster, M. A. Moran, and W. B. Whitman. 2011. Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. Nature 473:208-211.
- 60. Rinta-Kanto, J. M., H. Bürgmann, S. M. Gifford, S. Sun, S. Sharma, D. A. del Valle, R. P. Kiene, and M. A. Moran. 2011. Analysis of sulfur-related transcription by Roseobacter communities using a taxon-specific functional gene microarray. Environ. Microbiol. 13:453-467.
- Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, et al. 2007. The *Sorcerer II* global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. 5:e77.
- 62. Scarratt, M. G., M. Levasseur, S. Michaud, G. Cantin, M. Gosselin, and S. J.
  d. Mora. 2002. Influence of phytoplankton taxonomic profile on the distribution of dimethylsulfide and dimethylsulfoniopropionate in the northwest Atlantic. Mar. Ecol-Prog. Ser. 244:49-61.
- 63. Schäfer, H., N. Myronova, and R. Boden. 2010. Microbial degradation of dimethylsulphide and related C1-sulphur compounds: organisms and pathways controlling fluxes of sulphur in the biosphere. J. Exp. Bot. 61:315-334.
- Seymour, J. R., R. Simó, T. Ahmed, and R. Stocker. 2010. Chemoattraction to dimethylsulfoniopropionate throughout the marine microbial food web. Science 329:342-345.

- 65. Simó, R. 2001. Production of atmospheric sulfur by oceanic plankton:
  biogeochemical, ecological and evolutionary links. Trends Ecol. Evol. 16:287-294.
- 66. Simó, R., S. D. Archer, C. Pedros-Alio, L. Gilpin, and C. E. Stelfox-Widdicombe. 2002. Coupled dynamics of dimethylsulfoniopropionate and dimethylsulfide cycling and the microbial food web in surface waters of the North Atlantic. Limnol. Oceanogr. 47:53-61.
- Simó, R., and C. Pedros-Alio. 1999. Role of vertical mixing in controlling the oceanic production of dimethyl sulphide. Nature 402:396-399.
- Slezak, D., R. Kiene, D. Toole, R. Simó, and D. Kieber. 2007. Effects of solar radiation on the fate of dissolved DMSP and conversion to DMS in seawater. Aquat. Sci. 69:377-393.
- Stefels, J. 2000. Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. J. Sea Res. 43:183-197.
- Stefels, J., M. Steinke, S. Turner, G. Malin, and S. Belviso. 2007.
   Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. Biogeochemistry 83:245-275.
- Sunda, W., D. J. Kieber, R. P. Kiene, and S. Huntsman. 2002. An antioxidant function for DMSP and DMS in marine algae. Nature. 418:317-320.
- 72. **Taylor, B. F., and D. C. Gilchrist**. 1991. New routes for aerobic biodegradation of dimethylsulfoniopropionate. Appl. Environ. Microbiol. **57**:3581-3584.

- Todd, J. D., A. J. Curson, C. L. Dupont, P. Nicholson, and A. W. B. Johnston.
  2009. The *dddP* gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. Environ. Microbiol. 11:1624-1625.
- 74. Todd, J. D., A. R. J. Curson, M. Kirkwood, M. J. Sullivan, R. T. Green, and
  A. W. B. Johnston. 2011. DddQ, a novel, cupin-containing,
  dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine
  bacteria. Environ. Microbiol. 13:427-438.
- 75. Todd, J. D., M. Kirkwood, S. Newton-Payne, and A. W. B. Johnston. 2011. DddW, a third DMSP lyase in a model Roseobacter marine bacterium, *Ruegeria pomeroyi* DSS-3. ISME J. 6:223-226.
- 76. Todd, J. D., R. Rogers, Y. G. Li, M. Wexler, P. L. Bond, L. Sun, A. R. J. Curson, G. Malin, M. Steinke, and A. W. B. Johnston. 2007. Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. Science 315:666-669.
- Toole, D., D. J. Kieber, R. P. Kiene, D. A. Siegel, and N. B. Nelson. 2003.
  Photolysis and the dimethylsulfide (DMS) summer paradox in the Sargasso Sea.
  Limnol. Oceanogr. 48:1088-1100.
- Toole, D. A., and D. A. Siegel. 2004. Light-driven cycling of dimethylsulfide
   (DMS) in the Sargasso Sea: closing the loop. Geophys. Res. Lett. 31:L09308.

- 79. Tripp, H. J., J. B. Kitner, M. S. Schwalbach, J. W. H. Dacey, L. J. Wilhelm, and S. J. Giovannoni. 2008. SAR11 marine bacteria require exogenous reduced sulphur for growth. Nature 452:741-744.
- Turner, S. M., G. Malin, P. S. Liss, D. S. Harbour, and P. M. Holligan. 1988. The seasonal variation of dimethyl sulfide and dimethylsulfoniopropionate concentrations in nearshore waters. Limnol. Oceanogr. 33:364-375.
- 81. Vairavamurthy, A., M. O. Andreae, and R. L. Iverson. 1985. Biosynthesis of dimethylsulfide and dimethylpropiothetin by *Hymenomonas carterae* in relation to sulfur source and salinity variations. Limnol. Oceanogr. 30:59-70.
- Vallina, S. M., and R. Simó. 2007. Strong relationship between DMS and the solar radiation dose over the global surface ocean. Science 315:506-508.
- Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, et al.
   2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science
   304:66-74.
- 84. Vila-Costa, M., D. A. Del Valle, J. M. González, D. Slezak, R. P. Kiene, O. Sánchez, and R. Simó. 2006. Phylogenetic identification and metabolism of marine dimethylsulfide-consuming bacteria. Environ. Microbiol. 8:2189-2200.
- Vila-Costa, M., J. M. Rinta-Kanto, S. Sun, S. Sharma, R. Poretsky, and M.
   A. Moran. 2010. Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. ISME J. 4:1410-1420.
- Vila, M., R. Simó, R. P. Kiene, J. Pinhassi, J.M. González, M.A. Moran, C.
   Pedrós-Alió. 2004. Use of microautoradiography combined with fluorescence *in*

*situ* hybridization to determine dimethylsulfoniopropionate incorporation by marine bacterioplankton taxa. Appl. Environ. Microbiol. **70**:4648-4657.

- Wolfe, G. V., M. Steinke, and G. O. Kirst. 1997. Grazing-activated chemical defence in a unicellular marine alga. Nature 387:894-897.
- 88. Wolfe, G. V., S. L. Strom, J. L. Holmes, T. Radzio, and M. B. Olson. 2002. Dimethylsulfoniopropionate cleavage by marine phytoplankton in response to mechanical, chemical, or dark stress. J. Phycol. 38:948-960.
- 89. Yoch, D. C. 2002. Dimethylsulfoniopropionate: its sources, role in the marine food web, and biological degradation to dimethyl sulfide. Appl. Environ. Microbiol. 68:5804-5815.
- 90. Zubkov, M. V., B. M. Fuchs, S. D. Archer, R. P. Kiene, R. Amann, and P. H. Burkill. 2001. Linking the composition of bacterioplankton to rapid turnover of dissolved dimethylsulphoniopropionate in an algal bloom in the North Sea. Environ. Microbiol. 3:304-311.
- 91. Zubkov, M. V., B. M. Fuchs, S. D. Archer, R. P. Kiene, R. Amann, and P. H. Burkill. 2002. Rapid turnover of dissolved DMS and DMSP by defined bacterioplankton communities in the stratified euphotic zone of the North Sea. Deep-Sea Res. Pt. II. 49:3017-3038.

Fig. 1.1. Major DMSP degradation pathways in marine bacterioplankton: demethylation and cleavage pathways.



## CHAPTER 2

# DEEP SEQUENCING OF A DIMETHYLSULFONIOPROPIONATE DEGRADING GENE (DMDA) BY USING PCR PRIMERS DESIGNED ON THE BASIS OF MARINE METAGENOMIC DATA <sup>1</sup>

<sup>&</sup>lt;sup>1</sup> Varaljay, V.A., E.C. Howard, S. Sun, and M.A. Moran. 2010. Applied and Environmental Microbiology. 76: 609-617. Reprinted here with permission of publisher.

#### Abstract

In silico design and testing of environmental primer pairs with metagenomic data is beneficial for capturing a greater proportion of the natural sequence heterogeneity in microbial functional genes, as well as for understanding limitations of existing primer sets that were designed from more restricted sequence data. PCR primer pairs targeting ten environmental clades and subclades of the dimethylsulfoniopropionate (DMSP) demethylase protein, DmdA, were designed using an iterative bioinformatic approach that took advantage of thousands of *dmdA* sequences captured in marine metagenomic datasets. Using the bioinformatically-optimized primers, *dmdA* genes were amplified from composite free-living coastal bacterioplankton DNA (from 38 samples over 5 years and two locations) and sequenced using 454 technology. An average of 6,400 amplicons per primer pair represented more than 700 clusters of environmental *dmdA* sequences across all primers, with clusters defined conservatively at >90% nucleotide sequence identity (~95% amino acid identity). Degenerate and inosine-based primers did not perform better than specific primer pairs in retrieving *dmdA* richness, and sometimes captured a lower richness of sequences from the same DNA sample. A comparison of *dmdA* sequences in free-living versus particle-associated bacteria in southeastern U.S. coastal waters showed that sequence richness in some *dmdA* subgroups differed significantly between size fractions, though most gene clusters were shared (52-91%) and most sequences were affiliated with the shared clusters ( $\sim 90\%$ ). The availability of metagenomic sequence data has significantly enhanced the design of qPCR primer pairs for this key functional gene, providing robust access to the capabilities and activities of DMSP demethylating bacteria *in situ*.
## Introduction

Dimethylsulfoniopropionate (DMSP) is an abundant organic sulfur compound produced by marine phytoplankton as an osmolyte and for antioxidant purposes (5, 19, 27, 34, 36, 38). Upon cell lysis, DMSP and its degradation products are released into the surrounding seawater, thus providing bacterial communities with reduced organic carbon and sulfur (20) as well as contributing significantly to ocean-atmosphere sulfur flux (1, 1)24). Marine organisms capable of DMSP degradation can use either of two environmentally significant pathways. One route, known as the cleavage pathway, can lead to degassing of DMSP-derived sulfur from surface waters in the form of dimethylsulfide (DMS), an important catalyst in cloud formation. The second, a bacterial-specific route known as the demethylation pathway, results in DMSP-derived sulfur compounds [such as methylmercaptopropionate (MMPA) and methanethiol (MeSH)] that typically remain within the marine microbial food web. Studies show that certain groups of bacteria can mediate either or both competing pathways (11, 35), although the predominant route of DMSP degradation is through demethylation (18, 20, 21). Significant biogeochemical data for bacterially-mediated DMSP flux is now available (21, 33) and has allowed us to establish a framework for understanding this process in the marine environment (32). Yet the underlying genetic basis by which bacterioplankton perform and regulate these globally important sulfur transformations is relatively unknown.

The identification of dmdA (15), the gene encoding a DMSP demethylase that mediates the first step in the demethylation pathway, provides a key genetic tool for understanding DMSP fate in ocean waters. dmdA is highly abundant in marine

metagenomic datasets, with thousands of homologs (15, 16) identified in the Global Ocean Survey (GOS) Sargasso Sea (37) and 2007 GOS datasets (29). These findings indicate an important ecological role for *dmdA* in natural bacterioplankton communities. Two pressing areas for gene-based research include characterizing the diversity, abundance, and distribution of demethylating bacteria in the marine environment and determining how bacterial communities regulate DMSP fate via demethylation.

Here we describe our strategy for designing and testing *dmdA* primers to study the diversity of DMSP demethylating bacterial genes in marine environments. We took advantage of the non-PCR amplified *dmdA* homolog sequence reads identified in the 2007 GOS release to design universal and clade-specific primer pairs for *dmdA* sequences. An *in silico* primer testing pipeline checked specificity against metagenomic reads and identified mismatches to iteratively improve primer design. Primer pairs were tested empirically on free-living bacterial communities in nearshore waters of Sapelo Island, GA, U.S.A. using pyrosequencing to examine the deep diversity of *dmdA* amplicons. Selected primer pairs were then used to compare *dmdA* richness in gene reservoirs of the free-living and particle-associated communities.

## Materials and methods

#### Design of *dmdA* primer pairs

Metagenomic reads used in *dmdA* primer design were obtained from the Global Ocean Sampling (GOS) metagenome (29), with *dmdA* homologs in each of the five major clades (A through E) (Fig. 2.1) identified as previously described (16). DmdA sequences that were not in one of the major clades (11% of 1701 total sequences) were labeled as unclassified. These were used in primer design for the universal primer, but not the clade

or subclade primers. To identify subclades, the nucleotide sequences from the five major clades were clustered using MEGA version 3.1 (pairwise alignment, Jukes-Cantor algorithm, neighbor-joining model, 100 bootstrap replicates; (16) or Geneious Pro 3.5.6 (9) Tree-Builder (Tree global alignment: cost matrix 65% similarity (5.0/-4.0), gap open penalty 12, and gap extension penalty 3, with Jukes-Cantor algorithm, neighbor-joining model). Glycine cleavage T protein (gcvT) and related aminomethyl transferase (AMT) sequences served as outgroups. Subclades were defined as sequence clusters with bootstrap values  $\geq$  50% which captured at least 10% of reads in a clade. However not all subclades had conserved regions appropriate for primer design, and these could not be considered further (see below).

Subsets of *dmdA* nucleotide sequences were globally aligned in BioEdit Sequence Alignment Editor (14) and Geneious Pro 3.5.6 (9) programs using ClustalW. Primers were either designed manually or with the aid of Beacon Designer (Premier Biosoft International, Palo Alto, CA) primer design software. Primer pairs were designed to target amplicons without degeneracies ("specific" primer pairs), or included degenerate or inosine (a nucleoside that pairs indiscriminately) bases ("degenerate" and "inosine" versions) to accommodate common mismatches between primers and GOS reads that emerged from *in silico* testing (see below).

#### Bioinformatic pipeline: *in silico* primer tests

All primer pairs were iteratively tested *in silico* for specificity against the 1,701 *dmdA* sequences from the 2007 GOS release (Fig. 2.4 in Appendix A). Each GOS *dmdA* read was aligned to the *dmdA* gene from *Ruegeria pomeroyi* DSS-3 (SPO1913; 1,095 bp) to determine whether it contained the full region targeted by a given primer pair. Those

that did (designated "reads in range") were used for primer testing; those that did not were excluded. Primer pair specificity was then quantitatively assessed against GOS reads using an Exact Sequence and Pattern (ESP) Search program (http://web.chemistry.gatech.edu /~doyle/espsearch/) to determine the percent of reads successfully targeted by the primer pair. Sequences with mismatches were mined for number, location, and base of the mismatch. As a quality control check, the pipeline also determined if primers would bind non-specifically to sequences in non-target *dmdA* clades (including unclassified *dmdA* sequences).

A separate *in silico* test of non-specific binding of primers was also carried out against GOS metagenomic reads from three southeastern U.S. coastal sites (JCVI sites GS13, GS14, and GS15; ref. 29). All *dmdA* sequence reads were removed from these samples and the remaining 394,170 reads queried, allowing up to six total mismatches for forward plus reverse primers.

Primer pairs were either accepted or rejected based on results of the *in silico* testing, and if rejected were iteratively redesigned. Degenerate and inosine bases were incorporated into some finalized primers pairs if there were common mismatches, especially at a position away from the 5` end.

#### **DNA** samples

Surface water was collected between October 2000 and April 2005 at two sampling sites at the Sapelo Island Microbial Observatory (SIMO) (http://simo.marsci.uga.edu) in coastal Georgia, U.S.A. The Dean Creek site is a salt marsh tidal creek, and the Doboy Sound site is a coastal ocean inlet. To obtain each DNA sample, approximately 20 L of water were filtered sequentially through 8.0 µm, 1.0

 $\mu$ m, and 0.2  $\mu$ m polycarbonate membrane filters (Poretics Corp., Livermore, CA), with two replicate samples obtained at each location on each date. Cells captured on the 1.0  $\mu$ m filter (particle-associated) and the 0.2  $\mu$ m filter (free-living) were stored at -20 °C until DNA extraction with the PowerMax Soil DNA Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA). A total of 76 DNA extracts, representing 38 samples of each size fraction (free-living, 0.2 - 1.0  $\mu$ m; particle-associated, 1.0 - 8.0  $\mu$ m), were used in this study (Table 2.4 in Appendix A). These samples were separately pooled by size fraction in equal amounts to produce composite free-living and particle-associated DNA samples. Each composite sample encompassed temporal (seasonal/yearly) and spatial (tidal creek and coastal sound) variability at the SIMO site.

#### PCR amplicon preparation and sequencing

Primer pairs giving single amplicons of the correct size from the composite SIMO DNA were chosen for analysis by sequencing. Amplicons suitable for 454 sequencing were prepared by modifying each primer pair with an adaptor sequence at the 5` end of the forward primer according to Huber et al. (17). Additional 4-base key sequences in between the adaptor and primer sequence were used to distinguish inosine and degenerate primer sequences.

The typical PCR mix consisted of 0.5 U of Invitrogen (Carlsbad, CA) High Fidelity Platinum Taq polymerase, 0.2 mM dNTPs, and 2 mM MgSO4, although modifications of the MgSO4 concentrations were used for some primer pairs. Primer concentrations ranged from 0.2 to 0.8  $\mu$ M in final concentration in a 25.0  $\mu$ l reaction volume. PCR conditions were as follows: initial denaturing at 94 °C for 2.0 min, 30-40 cycles of denaturing at 94 °C for 20 s, annealing at various temperatures (see Table 2.1)

for 30 s, extension at 68 °C for 30 s and a final extension at 68 °C for 5.0 min. All PCR reactions were carried out in duplicate using 24 ng template DNA and then pooled before sequencing. For the Clade C/2 inosine primer pair, four PCR reactions were pooled because of low amplicon abundance. Pooled products were cleaned (QIAGEN, Valencia, CA, QIAquick PCR Purification kit) and stored at -20 °C; for some products, an additional gel excision step was included (QIAGEN, Valencia, CA, QIAquick Gel Extraction kit). Amplicons were cleaned using the AMPure purification method (Agencourt Bioscience Corp., Beverly, MA) according to the 454 Life Sciences protocol (Roche Diagnostics Corp., Branford, CT), with modifications to the volume of purified PCR products (30.0 μl) and AMPure beads (50.4 μl). Products were quantified spectrophotometrically and combined in equal concentrations in four separate pools based on primer and size fraction. Four-region 454 FLX LR70 sequencing was carried out at the University of South Carolina EnGenCore facility.

#### **Clustering and clade designations**

After removal of low quality reads (< 20 quality score;  $\leq 0.03\%$  of sequences), primer sequences were stripped from the remaining 252,319 reads. For the universal primer pair, sequence data was obtained for the first ~250 bases. For the other primer pairs, the full amplicon was sequenced. Within a primer pair (including specific, degenerate, and inosine versions when applicable) sequences were clustered based on 90% nucleotide identity (Cd-hit clustering; ref. 23). Given an error rate for 454 sequencing of 0.3% (25), sequencing errors should not change cluster assignments, but would inflate estimates of unique sequences.

Amplicon sequences were annotated by BLASTx analysis using a default maximum E-value of 10 against an in-house database which consisted of DmdA and related non-DmdA sequences from the GOS metagenome and cultured organisms. This analysis was used to distinguish correct target sequences from closely-related paralogous sequences and to classify amplicons by clade. The high E-value cut-off reflected the short length of the query sequences (e.g., 39 bp for the Clade D/1 amplicons after primer sequences were stripped), but most hits had percent similarities of >90%. The BLAST database consisted of 3,280 total protein sequences (assembled from the Sargasso Sea GOS dataset, the 2007 GOS dataset, the Indian Ocean GOS dataset, and cultured organisms; refs. 15, 16 and http://camera.calit2.net), including sequences from Clade A (146 sequences), Clade B (76), Clade C (407), Clade D (1792), and Clade E (19), as well as unclassified DmdA sequences (217), and non-target *gcvT* and aminomethyltransferase sequences (623). Of the 3,280 sequences in the database, ~20 were DmdA sequences from cultured organisms.

#### **Richness and shared sequence analyses**

To account for differences in the number of amplicons sequenced for each primer pair (ranging from 2,000 to 12,000 sequences), a resampling approach was used in which 1000 sample populations of the same size were randomly drawn from the amplicon pools being compared. This approach was used to normalize the number of 90% *dmdA* clusters in comparisons between primer pairs and size fractions. Statistical significance was assigned based on the distribution of pairwise differences between the 1000 random populations using a 95% confidence interval (12). Rarefaction curves for a primer pair was based on 90% sequence clusters using EcoSim 7.0 (13) with 1,000 resamplings.

## Nucleotide sequence accession numbers

The nucleotide sequences of *dmdA* 454-sequenced PCR amplicons were deposited in the GenBank Short Read Archive (SRA) under the accession number SRA008804.8. **Results** 

#### in silico dmdA primer design

The 1,701 *dmdA* sequences identified from the 2007 GOS metagenome (16) served as the database for designing hierarchical PCR primer pairs for the DMSP demethylase gene (Fig. 2.1). Primer design efforts focused on a universal primer pair, to capture as many *dmdA* sequences as possible from marine environmental samples, as well as on clade and subclade primer pairs to capture conserved sequence subsets within the five known clades of *dmdA*. Multiple alignments of a subset of target sequences (up to 50) were used for initial primer design. We avoided AT rich regions (particularly problematic for Clades C and D), long nucleotide repeats, sequences that might lead to primer dimers, and regions with high similarity to glycine cleavage T genes or other related non-*dmdA* genes. Primer pairs were tested *in silico* against the remaining sequences, followed by design optimization to complement the greatest number of identified *dmdA* sequences. The pipeline (Fig. 2.4 in Appendix A) simultaneously checked for matches to non-target sequences, including sequences in the incorrect dmdA clade or subclade, or sequences of paralogous genes (i.e., gcvT and related aminomethyltransferases; Fig. 2.1).

While the original goal was to design all primers for use in qPCR, sufficiently conserved primer areas flanking a small (<250 bp) region of the gene could not be identified for a universal primer pair. However, a universal *dmdA* primer pair amplifying

a larger region (537 bp) from sequences in all five protein clades and targeting >90% of 2007 GOS *dmdA* reads in range with <2 mismatches per primer when degeneracies were included was identified (Table 2.1).

A clade-specific qPCR primer pair was designed for Clade D; Clades A, B, C, and E were highly diverse at the nucleotide level and primers were targeted instead to the abundant subclades (Table 2.1 and Fig. 2.1). Although smaller subsets of diverse sequences were not considered in primer design with this approach, they accounted for only ~20% of the 1,701 GOS *dmdA* sequences. In order to accommodate as many sequences as possible, clade and subclade primer pairs were designed without degeneracies ("specific" primer pairs), or included degenerate or inosine (a nucleoside that pairs indiscriminately) bases ("degenerate" and "inosine" versions) to accommodate common mismatches. When primer design was completed, the clade and subclade primer pairs targeted an average of 70% (with  $\leq$ 4 mismatches) or 80% (with  $\leq$ 6 mismatches) of *dmdA* reads in the correct clade (Table 2.1 and Appendix A Table 2.5), although success rate was as low as 20% for one primer pair. Subclade C/1 and D/2 primers targeted few sequences based on results of the bioinformatic analyses, and were not considered further.

An *in silico* check for non-specific primer binding was carried out against non*dmdA* metagenomic reads from three coastal sites in the 2007 GOS (sites GS13, GS14, and GS15; ref. 29); these were selected because they are geographically closest to the source of environmental DNA used in this study (see below). Less than 150 of the ~350,000 non-*dmdA* metagenomic reads were complementary to both primers in any pair, even with an allowance of six mismatches per primer pair, and none of these would

produce an amplicon of the correct size. Overall, final primer designs from the bioinformatic pipeline resulted in twenty two primer pairs (which included degenerate and inosine versions where applicable) to fourteen target groups: one universal target group, one clade-specific target group (Clade D), and 12 subclade-specific target groups (three in Clade A, four in Clade B, one in Clade C, two in Clade D, and two in Clade E; Table 2.1).

## **Experimental primer testing**

All *in silico*-tested primer pairs (including degenerate and inosine versions) were tested experimentally using composite DNA from free-living bacterioplankton communities ( $0.2 - 1.0 \mu m$  size fraction) collected over 5 years at the Sapelo Island Microbial Observatory (SIMO; http://simo.marsci.uga.edu) (Table 2.4 in Appendix A). DNA from 38 different samples was combined in order to capture the temporal and spatial variability of *dmdA* sequences at this coastal site, while keeping the number of amplicon pools to a reasonable level for sequencing. Of the 14 target groups, *dmdA* primer pairs to four (A/3, B/1, B/2, and E/1) did not produce amplicons from the composite DNA samples. Since these primers passed all bioinformatic criteria, they are described in the supplementary material (Table 2.5 in Appendix A) for potential use in PCR-based analyses of *dmdA* in other marine environments. The remaining ten groups were targeted by eighteen primer pairs (including degenerate and inosine versions; Table 2.1) that successfully produced amplicons from the composite DNA sample.

Amplicons were sequenced using 454 pyrosequencing technology and annotated based on the best hit in BLASTx analysis against our 3,280-member *dmdA* database (Table 2.2). An *in silico* test of known *dmdA* sequences with priming sites trimmed

indicated that the BLAST analysis was accurate in assigning sequences to clades despite the short amplicons produced by some primer pairs (e.g., Clade D/1 primers produce a 39 bp trimmed amplicon). For each primer pair, we determined: 1) the percent correct sequences retrieved by the primers (as opposed to sequences with best hits to the wrong clade or to *dmdA* paralogs, or that had no hit; some of these might include novel *dmdA* genes); and 2) the richness of *dmdA* sequence clusters retrieved by the primers, defining clusters at a >90% nucleotide (~95 % amino acid) identity level and using a resampling approach to normalize for differences in number of sequences between primer pairs (see Methods).

For the universal primer pair, the majority of the sequences were *dmdA* (94%), with only a small number having better homology to paralogous genes or having no hits in the BLAST analysis (6%) (see Table 2.1; two different annealing temperatures were tested for the universal primer pair, but both yielded similar numbers of correct *dmdA* sequences). Cluster analysis indicated 116 *dmdA* clusters were retrieved from the composite free-living bacterioplankton DNA and these sequences represented all five major clades (Fig. 2.2). Clades A and D amplicons were the most abundant in terms both of numbers of sequences and numbers of clusters (Fig. 2.2)

For most specific clade and subclade primer pairs, at least 90% of sequences were *dmdA* from the correct target clade (Table 2.2). The majority of non-specific hits were to unclassified *dmdA* sequences, and less than 1% of hits were to paralogous proteins. For most subclade primers, ~98% of amplicons hitting the correct clade also hit the correct subclade (Fig. 2.2). Summing across all specific primer pairs for the targeted clades and subclades, cluster analysis indicated that 600 total clusters and up to 17,203 unique

nucleotide sequences were retrieved (from a total of 62,606 sequences). *dmdA* richness cannot be compared between clades or subclades using these primer pairs, however, because the regions of the gene targeted by the primers differ.

## Specific versus degenerate primer pairs

Primer pairs with degenerate or inosine positions were included for some target groups if the bioinformatic pipeline indicated they might substantially improve retrieval of *dmdA* diversity. The degenerate/inosine primer pairs were no more likely to retrieve incorrect sequences than the specific primers (Fig. 2.2), indicating that the modifications did not cause undue problems with non-specific amplification. However, they were also no more likely to retrieve a higher richness of *dmdA* sequences than the specific primers (as defined by 90% nucleotide sequence clusters) (Table 2.2; except for Clade C/2 degenerate primers). Moreover, most of the *dmdA* sequences retrieved with modified primers were the same as those retrieved with the specific primers (Fig. 2.5 in Appendix A), and a similar percent of unique clusters were captured with the modified and specific primers. Thus, for this particular functional gene, primers modified with degenerate or inosine bases did not retrieve a richer sequence library. Based on similar performance of these primer types and potential complications of using modified primers in future qPCR applications, only amplicons of the specific versions of the primer pairs were used in a subsequent comparative analysis of free-living versus particle-associated bacterioplankton communities.

#### *dmdA* in free-living and particle-associated bacterial communities.

The *dmdA* sequences amplified with the universal primer pair from southeastern U.S. coastal waters had comparable clade distributions in both the particle-associated (1.0

- 8.0  $\mu$ m) and free-living (0.2 - 1.0  $\mu$ m) size fractions. Clades A and D made up the majority of sequences in both fractions (Fig. 2.2A and Appendix A Table 2.6 footnote). The universal primer pair targeted a higher number of apparent non-*dmdA* sequences in the particle-associated fraction (19%) compared to the free-living fraction (6%) (Tables 2.2 and Appendix A Table 2.6), but also showed a higher richness of correct *dmdA* clusters in the particle-associated community (Fig. 2.3). The clusters shared between the two communities accounted for most of the sequences (91%), and unique clusters were small in size (~2 sequences per cluster).

Amplicon richness and composition for clades and subclades of *dmdA* retrieved with the specific primer pairs were also comparable for free living and particle-associated bacteria (Table 2.6 in Appendix A). An average of 60% of the clusters were shared across size fractions (Table 2.3). While four of nine subclade primer pairs showed a significant difference in the number of unique clusters retrieved between size fractions (Table 2.3), in some cases richness was higher in the particle associated fraction (Clade C/2), and in some cases in the free-living fraction (Clades A/2, B/3, and D/1) (Fig. 2.3). However, unique clusters typically had few sequences and, like the universal primer pair, an average of 90% of *dmdA* sequences obtained with clade and subclade primer pairs were members of clusters shared across the size fractions.

### Discussion

The advent of metagenomic sequencing offers a significant advantage in environmental primer design. Previously, sequences from cultured organisms or small environmental clone libraries formed the basis for primer sequences. Yet how well those primers targeted the full natural gene diversity, and therefore captured gene abundance,

distribution, and expression in complex bacterial communities (3, 31), was not known. Ecologically relevant sequences from metagenomic data are now available for designing primers for field studies (6). Here we made use of the thousands of *dmdA* homologs from marine metagenomic data to design optimized primer pairs, and then systematically assessed the primers by deep sequencing of amplicon populations. The substantial nucleotide sequence diversity in the GOS dataset for this single gene made it necessary to target groups at the subclade level. Similarly high levels of richness have been found for another widespread and abundant marine bacterial gene, proteorhodopsin (3).

When primers were tested on coastal DNA, more than 90% of the amplicons were from the correct *dmdA* target group. The universal primer pair captured all five clades, with a significant proportion of correct sequences classified as Clade A (43%) or D (37%). These two clades harbor genes from cultured roseobacters and SAR11 members, respectively, and were also abundant among *dmdA* genes retrieved from coastal and open ocean sites in the GOS dataset (16). Other primer pairs for clades and subclades of *dmdA* were likewise highly specific in targeting correct sequences. Overall, the *dmdA* amplicons formed hundreds of clusters at >90% nucleotide (~95% amino acid identity, based on manual alignments of translated sequences from a subset of clusters) and did not reach full saturation even after ~6,400 sequences per primer pair. Since a composite DNA preparation from 38 samples was used to assess primer performance (to increase the likelihood of target genes for each primer pair being tested), we do not yet know how abundance and composition of the *dmdA* pool varies over time and space; these vetted qPCR primers now provide a robust tool to address *dmdA* dynamics in this and other locations.

The modification of primer sequences with degenerate bases or inosine has been used previously to improve PCR primer annealing when target sequences are heterogeneous (3, 10, 22, 31, 39). For environmental primers, such modifications might allow more of the natural diversity of a functional protein to be captured (39), although potentially at the expense of non-specific binding. In this study, modified primers were no more prone to non-specific amplification than specific primers. Yet while we expected that amplicons from the unmodified parent primers would be a subset of those from the modified primers, surprisingly this was not the case for this study. Generalizing across the primer pairs tested, the degenerate and inosine primers captured an equally diverse but slightly different suite of sequences compared to the specific primers. These empirical results guided us toward the use of the specific clade and subclade primers in subsequent analyses. We did not design or test a specific version of the universal *dmdA* primer.

In the first use of these primer pairs, we asked whether the composition of the *dmdA* reservoir (based on 38 pooled samples spanning 5 years) differs between freeliving and particle-attached bacterial communities in southeastern U.S. coastal waters. The GOS metagenomic dataset, which comprises the largest collection of environmental *dmdA* sequences to date, is heavily biased toward free-living cells (defined as <0.8  $\mu$ m diameter), providing little information on representation of the major clades and subclades of *dmdA* in particle-associated communities. DMSP concentrations are locally higher in marine particle "microenvironments" than in bulk seawater (20), since the primary source of DMSP is phytoplankton cells, raising the issue of whether particle-associated demethylation is driven by a different suite of *dmdA* orthologs. Differences in

*dmdA* composition between the two size fractions could reflect ecological advantages conferred by differing kinetic parameters of the major clades (e.g., Km and Kcat; ref. 28). Alternatively, taxonomic differences between marine bacterial size classes, as has been shown previously (8), may drive differences in the composition of the *dmdA* reservoirs. In either case, gene composition might provide insights into rates of, or controls on, DMSP demethylation. DMSP lyase activity (i.e., the competing pathway for DMSP degradation) has been shown to be greater in particle-associated microbial communities than in free-living (4, 30).

Here, we used a 1.0 µm pore size filter to operationally separate free-living from particle-associated bacteria, and conducted a comparative analysis of their *dmdA* reservoirs. While the universal primer pair suggested higher overall sequence richness in the particle-attached communities (Table 2.3), results were mixed for individual clade and subclade primer pairs: one primer pair also retrieved significantly higher richness in the particle-associated fraction; three retrieved significantly higher richness in the freeliving fraction; and five showed no difference. Since Clade D primers likely target *dmdA* sequences in SAR11 populations (15), we predicted higher richness for this clade of planktonic oligotrophs (26) in the free-living size fraction, and this was the case (Fig. 2.3). Since Clade A primers target *dmdA* sequences in Roseobacter cells (and other taxa), we predicted higher richness for this clade of surface colonizers (2, 7) in the particleassociated fraction, but this was not the case. Yet despite these significant differences in cluster richness for some primer pairs, the vast majority of sequences were assigned to clusters that were shared between free-living and particle-attached cells (Table 2.3). Since our primers were designed from the GOS metagenome, which mostly includes

free-living bacterioplankton in the 0.2 - 0.8  $\mu$ m size range, we cannot rule out the possibility that they systematically miss *dmdA* diversity in particle-associated bacteria. Better metagenomic coverage of larger size classes of marine particles in future sequencing efforts will provide a mechanism to check, and if necessary redesign, *dmdA* primers.

The availability of metagenomic sequence data has greatly improved our ability to design qPCR primers to assess abundance, diversity, and expression of microbial functional genes in the environment. In the case of the DMSP demethylase, knowledge of how *dmdA* genes vary over time and space, and how their expression changes in response to DMSP dynamics and environmental drivers, will increase understanding of the marine bacterial communities that regulate sulfur emission from the ocean surface.

## Acknowledgements

We thank W. Ye and C. Lasher for DNA samples; S. Sharma for bioinformatic expertise; C. English for graphics assistance; R. Newton, W. Whitman, A. Karls, S. Gifford, and J. Edmonds for helpful discussions; and J. Jones at the University of South Carolina EnGenCore Sequencing Facility for sequencing expertise. This research was supported by grants from the National Science Foundation (OCE-0724017) and the Gordon and Betty Moore Foundation.

### References

- Andreae, M. O. 1990. Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. Mar. Chem. 30:1-29.
- Buchan, A., J. M. Gonzalez, and M. A. Moran. 2005. Overview of the marine Roseobacter lineage. Appl. Environ. Microbiol.71:5665-5677.
- Campbell, B. J., L. A. Waidner, M. T. Cottrell, and D. L. Kirchman. 2008.
  Abundant proteorhodopsin genes in the North Atlantic Ocean. Environ.
  Microbiol. 10:99-109.
- Cantin, G., M. Levasseur, S. Schultes, and S. Michaud. 1999. Dimethylsulfide (DMS) production by size-fractionated particles in the Labrador Sea. Aquat. Microb. Ecol. 19:307-312.
- Chambers, S. T., C. M. Kunin, D. Miller, and A. Hamada. 1987.
  Dimethylthetin can substitute for glycine betaine as an osmoprotectant molecule for *Escherichia coli*. J. Bacteriol. 169:4845-4847.
- Contreras-Moreira, B., B. Sachman-Ruiz, I. Figueroa-Palacios, and P. Vinuesa. 2009. Primers4clades: a web server that uses phylogenetic trees to design lineage-specific PCR primers for metagenomic and diversity studies. Nucleic Acids Res. 37:W95–W100. doi:10.1093/nar/gkp377.
- Dang, H., and C. R. Lovell. 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. Appl. Environ. Microbiol. 66:467-475.

- DeLong, E. F., D. G. Franks, and A. L. Alldredge. 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol. Oceanogr. 38:924-934.
- Drummond, A., B. Ashton, M. Cheung, J. Heled, M. Kearse, R. Moir, S. Stones-Havas, T. Thierer, and A. Wilson. 2008. Geneious, v4.0. Biomatters Ltd., Auckland, New Zealand. 2008. Geneious v4.0.
- Ehlen, T., and L. Dubeau. 1989. Detection of ras point mutations by polymerase chain reaction using mutation-specific, inosine-containing oligonucleotide primers. Biochem. Biophys. Res. Commun. 160:441-447.
- Gonzalez, J. M., R. P. Kiene, and M. A. Moran. 1999. Transformation of sulfur compounds by an abundant lineage of marine bacteria in the α-subclass of the class *Proteobacteria*. Appl. Environ. Microbiol. 65:3810-3819.
- Good, P. I. 2005. Introduction to statistics through resampling methods and R/S-PLUS, vol. John Wiley and Sons, Inc., Hoboken, NJ.
- Gotelli, N. J., and G. L. Entsminger. 2008. EcoSim: Null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear, Jericho.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids. Symp. Ser. 41:95-98.
- Howard, E. C., J. R. Henriksen, A. Buchan, C. R. Reisch, H. Buergmann, et al. 2006. Bacterial taxa that limit sulfur flux from the ocean. Science 314:649-652.

- Howard, E. C., S. Sun, E. J. Biers, and M. A. Moran. 2008. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. Environ. Microbiol. 10:2397-2410.
- Huber, J. A., D. B. M. Welch, H. G. Morrison, S. M. Huse, P. R. Neal, D. A.
  Butterfield, and M. L. Sogin. 2007. Microbial population structures in the deep marine biosphere. Science 318:97-100.
- Kiene, R. P. 1996. Production of methanethiol from dimethylsulfoniopropionate in marine surface waters. Mar. Chem. 54:69-83.
- Kiene, R. P., and L. J. Linn. 2000. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico Limnol. Oceanogr. 45:849-861.
- Kiene, R. P., L. J. Linn, and J. A. Bruton. 2000. New and important roles for DMSP in marine microbial communities. J. Sea Res. 43:209-224.
- Kiene, R. P., L. J. Linn, J. Gonzalez, M. A. Moran, and J. A. Bruton. 1999. Dimethylsulfoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. Appl. Environ. Microbiol. 65:4549-4558.
- 22. Knoth, K., S. Roberds, C. Poteet, and M. Tamkun. 1988. Highly degenerate, inosine-containing primers specifically amplify rare cDNA using the polymerase chain reaction. Nucl. Acids Res. 16:10932.
- Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22:1658-1659.

- 24. Lovelock, J. E., R. J. Maggs, and R. A. Rasmussen. 1972. Atmospheric dimethyl sulphide and the natural sulphur cycle. Nature 237:452-453.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.
- 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.
- 27. Pichereau, V., J.-A. Pocard, J. Hamelin, C. Blanco, and T. Bernard. 1998. Differential effects of dimethylsulfoniopropionate, dimethylsulfonioacetate, and other s-methylated compounds on the growth of *Sinorhizobium meliloti* at low and high osmolarities. Appl. Environ. Microbiol. **64**:1420-1429.
- 28. **Reisch, C. R., M. A. Moran, and W. B. Whitman**. 2008.

Dimethylsulfoniopropionate-dependent demethylase (DmdA) from *Pelagibacter ubique* and *Silicibacter pomeroyi*. J. Bacteriol. **190**:8018-8024.

- Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, et al. 2007. The *Sorcerer II* Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. 5:e77.
- 30. Scarratt, M., G. Cantin, M. Levasseur, and S. Michaud. 2000. Particle sizefractionated kinetics of DMS production: where does DMSP cleavage occur at the microscale? J. of Sea Res. 43:245-252.
- 31. Schwalbach, M. S., and J. A. Fuhrman. 2005. Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by

epifluorescence microscopy and quantitative PCR. Limnol. Oceanogr. **50**:620-628.

- Simo, R. 2004. From cells to globe: approaching the dynamics of DMS(P) in the ocean at multiple scales. Can. J. Fish. Aquat. Sci. 61:673-684.
- Simo, R. 2001. Production of atmospheric sulfur by oceanic plankton:
  biogeochemical, ecological and evolutionary links. Trends Ecol. Evol. 16:287-294.
- Sunda, W., D. J. Kieber, R. P. Kiene, and S. Huntsman. 2002. An antioxidant function for DMSP and DMS in marine algae. Nature 418:317-320.
- Taylor, B. F., and D. C. Gilchrist. 1991. New routes for aerobic biodegradation of dimethylsulfoniopropionate. Appl. Environ. Microbiol. 57:3581-3584.
- 36. Vairavamurthy, A., M. O. Andreae, and R. L. Iverson. 1985. Biosynthesis of dimethylsulfide and dimethylpropiothetin by *Hymenomonas carterae* in relation to sulfur source and salinity variations. Limnol. Oceanogr. **30**:59-70.
- Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, et al.
  2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science
  304:66-74.
- Yoch, D. C. 2002. Dimethylsulfoniopropionate: Its sources, role in the marine food web, and biological degradation to dimethylsulfide. Appl. Environ. Microbiol. 68:5804-5815.
- Yutin, N., M. T. Suzuki, and O. Beja. 2005. Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. Appl. Environ. Microbiol. 71:8958-8962.

Table 2.1. Eighteen *dmdA* primer pairs (including degenerate and inosine versions) targeting ten sequence groups and results of *in silico* testing against the 2007 GOS dataset.

Primer name	Primer version	<i>dmdA</i> position <sup>a</sup>	Amplicon length (bp)	Amplicon Primer sequence <sup>b</sup> A length (bp)		No. target GOS reads	No. target GOS reads	No. (%) rea binding $\leq 4$ mismatches	ads in range primers $\leq 6$ mismatches
							in range <sup>d</sup>		
dmdAU	N/A	157-694	537	dmdAUF160 – GTICARITITGGGAYGT dmdAUR697 - TCIATICKITCIATIAIRTTDGG	32 and 41°	1701	1041, 1093	993, 991 (93%)	N/A
A/1	specific	368-596	228	A/1-spFP - ATGGTGATTTGCTTCAGTTTCT A/1-spRP – CCCTGCTTTGACCAACC	53	30	16	13 (81%)	15 (94%)
A/2	specific	339-486	147	A/2-spFP - CGATGAACATTGGTGGGTTTCTA A/2-spRP - GCCATTAGGTCGTCTGATTTTGG	59	16	10	4 (40 %)	7 (70%)
	degener ate	339-486	147	A/2-dgFP - YGATGAWCATTGGTGGGTTTCKA A/2-dgRP - GCCATYARGTCGTCYGATTTTGG	58	16	10	8 (80%)	9 (90%)
	inosine	339-486	147	A/2-inoFP - IGATGAICATTGGTGGGTTTCIA A/2-inoRP - GCCATIAIGTCGTCIGATTTTGG	57	16	10	8 (80%)	9 (90%)
B/3	specific	169-323	154	B/3-spFP - GATGTCTCCTGCCAACGTCAGGTCGA B/3-spRP - ACCGGGTCATTGATCATGCCTGCG	62	4	3	3 (100%)	3 (100%)

B/4	specific	361-553	192	B/4-spFP – ATTGCCGACTCGGATGTTCT B/4-spRP - CAAGAAGGTCAAACATGGCAAAC	58	5	4	4 (100%)	4 (100%)
C/2	specific	291-482	191	C/2-spFP - AGATGAAAATGCTGGAATGATAAATG C/2-spRP - AAATCTTCAGACTTTGGACCTTG	50	141	94	19 (20%)	44 (47%)
	degener ate	291-482	191	C/2-dgFP - AGATGAAAATGCWGGRATGATAAATG C/2-dg RP - AAWTCTTCAGAYTTTGGACCTTG	52	141	94	44 (47%)	55 (60%)
	inosine	291-482	191	C/2-inoFP - AGATGAAAATGCIGGIATGATAAATG C/2-inoRP - AAITCTTCAGAITTTGGACCTTG	52	141	94	44 (47%)	57 (61%)
D/1	specific	268-356	89	D/1-spFP - AGATGTTATTATTGTCCAATAATTGATG D/1-spRP - ATCCACCATCTATCTTCAGCTA	49	402	268	110 (41%)	189 (71%)
D/3	specific	347-473	126	D/3-spFP - AATGGTGGATTTCTATTGCAGATAC D/3-spRP GATTTTGGACCTTGTACAGCCA	54	262	155	94 (61%)	116 (75%)
	degener ate	347-473	126	D/3-dgFP - AATGGTGGRTTTCTATTGCWGATWC D/3-dgRP - GATTTWGGMCCTTGYACAGCCA	56	262	155	113 (73%)	137 (88%)
D/all	specific	984- 1089	105	D/all-spFP – TATTGGTATAGCTATGAT D/all-spRP – TAAATAAAAGGTAAATCGC	42	1125	457	190 (42%)	320 (70%)
	degener ate	984- 1089	105	D/all-dgFP – TATTGGTATWGCWATGAT D/all-dgRP – TAAATRAAAGGYAAATCGC	41	1125	457	324 (71%)	394 (86%)

	inosine	984- 1089	105	D/all-inoFP – TATTGGTATIGCIATGAT D/all-inoRP – TAAATIAAAGGIAAATCGC	48	1125	457	346 (76%)	417 (91%)
E/2	specific	80-154	133	E/2-spFP – CATGTTCAGATCTGGGACGT E/2-spRP - AGCGGCACATACATGCACT	57	4	2	2 (100%)	2 (100%)
	degener ate	80-154	133	E/2-dgFP – CATGTTCAGATMTGGGAYGT E/2-dgRP – AGCGGCAYATACATGCACT	56	4	2	2 (100%)	2 (100%)

<sup>a</sup>Position numbers based on the full length *dmdA* sequence in *Ruegeria pomeroyi* DSS-3 (SPO1913).

<sup>b</sup>Degenerate codes are as follows:  $R = \tilde{A}$  or G,  $Y = \tilde{C}$  or T, W = A or T, M = A or C, K = G or T.

<sup>°</sup>Two annealing temperatures were used in separate PCR reactions.

<sup>d</sup> "Reads in range" refers to sequences that span the full region between the forward and reverse primers, allowing both to be tested for complementarity. In the case of the universal primer pair, the larger amplicon size required that the forward and reverse primers be tested with different subsets of reads, resulting in different numbers of reads in range for each primer.

Primer name	Clade	Subclade	% with correct clade(s) targeted	% with correct subclade (of correct clade targeted)	% with incorrect clade targeted	% not <i>dmdA</i> <sup>b</sup>	No. sequences resampled	Normalized no. <i>dmdA</i> clusters <sup>c</sup>
dmdAU	All	All	94.0	n/a	n/a	6.0	400	51
A/1-sp	Clade A	Subclade 1	99.2	99.8	0.5	0.3	2500	30
A/2-sp	Clade A	Subclade 2	98.7	97.8	0.3	1.0	3500	25
A/2-dg	Clade A	Subclade 2	99.1	99.4	0.1	0.8	3500	24
A/2-ino	Clade A	Subclade 2	99.4	99.4	0.05	0.5	3500	20*
B/3-sp	Clade B	Subclade 3	97.6	97.9	1.5	0.9	5500	46
B/4-sp	Clade B	Subclade 4	33.6	99.3	65.4	0.9	1500	20
C/2-sp	Clade C	Subclade 2	92.5	68.8	6.3	1.2	1200	23
C/2-dg	Clade C	Subclade 2	64.2	81.8	33.8	2.0	1200	35*
C/2-ino	Clade C	Subclade 2	71.7	98.8	26.2	2.2	1200	20
D/1-sp	Clade D	Subclade 1	88.4	97.8	0.5	11.2	6000	200
D/3-sp	Clade D	Subclade 3	99.6	91.5	0.10	0.3	4300	30
D/3-dg	Clade D	Subclade 3	95.3	96.7	4.6	0.1	4300	32
D/all-sp	Clade D	All	99.3 <sup>d</sup>	n/a	0.2	0.5	4500	82
D/all-dg	Clade D	All	99.9 <sup>d</sup>	n/a	0.1	0	4500	68*
D/all-ino	Clade D	All	99.8 <sup>d</sup>	n/a	0	0.2	4500	74*
E/2-sp	Clade E	Subclade 2	96.65	99.99	1.58	1.77	3000	43
E/2-dg	Clade E	Subclade 2	98.97	99.78	0.51	0.51	3000	35*

Table 2.2. BLASTx and clustering results for *dmdA* amplicons of the free-living size fraction from southeastern U.S. coastal seawater<sup>a</sup>

<sup>a</sup> For particle-associated data, see Table 2.6 in Appendix A. n/a, not applicable.

<sup>b</sup>Includes sequences with hits to gcvT and those with no hits. <sup>c</sup>Average of 1000 resamplings (see Methods) using the population sizes indicated in the "No. sequences resampled" column. Cluster numbers marked with an asterisk were significantly different (p < 0.05) from that obtained by the specific version of that primer pair. <sup>d</sup> For the D/all-sp primer pair, 16.2% of hits were to Subclade D/1 and 2.5% to Subclade D/3; for the D/all-dg primer pair, 13.8% of hits were to Subclade D/1 and 6.7% to Subclade D/3; and for the D/all-

ino primer pair, 4.38% of hits were to Subclade D/1 and 6.3% to Subclade D/3. Remaining correct hits were to Clade D sequences not classified within a subclade.

	% unique	clusters <sup>a</sup>	% shared clusters	% sequences in unique clusters		% sequences in shared clusters
	Free-living	Particle associated	-	Free-living	Particle associated	
Univ	24	43*	33	3	6	91
A/1-sp	7	15	78	<1	<1	99
A/2-sp	33*	10	57	7	<1	93
B/3-sp	31*	5	64	27	<1	73
B/4-sp	10	6	84	<1	<1	>99
C/2-sp	23	33*	44	2	<1	98
D/1-sp	29*	17	54	21	<1	79
D/3-sp	19	15	66	<1	<1	>99
D/all-sp	21	22	57	11	<1	88
E/2-sp	15	13	72	5	<1	95

Table 2.3. Unique and shared clusters and percent shared sequences between size fractions for ten *dmdA* primer pairs.

<sup>a</sup> Average of 1000 resamplings (see Methods) using population sizes indicated in Table 2.2. Cluster numbers marked with an asterisk were significantly higher (p < 0.05) from those obtained for the other size fraction.

Figure 2.1. Amino acid tree of representative GOS DmdA sequences. The wedge size is approximately proportional to the number of sequences within the group. Selected DmdA homologs from cultured marine bacteria are included. 'Additional cultured Roseobacters' include *Roseobacter denitrificans* Och114, *Roseobacter* sp. Azwk3b, *Roseobacter* sp. MED193, *Roseovarius* sp. 217, *Roseovarius nubinhibens* ISM, *Roseovarius* sp. TM1035, and *Ruegeria* sp. TM1040.Related glycine cleavage T (gcvT) and aminomethyltransferase (AMT) sequences serve as outgroups. Bootstrap values of <50 have been removed for clarity. The neighbor-joining tree was made with Jones-Taylor-Thornton distances. The exact position of the cluster designated Clade C/1 can vary depending on the sequences included in the tree (data not shown).



Figure 2.2. Annotation of free-living  $(0.2 - 1.0 \,\mu\text{m})$  amplicon sequences from *dmdA* primer pairs based on best hits in a BLASTx analysis against known *dmdA* sequences. A) Universal primer pair. B) Clade and subclade primer pairs, including specific, degenerate, and inosine versions.



Figure 2.3. Rarefaction curves of *dmdA* amplicons from free-living and particleassociated bacterioplankton communities based on 90% nucleotide identity clusters. A) Universal *dmdA* primer pair. B) Selected subclade primer pairs.



## CHAPTER 3

# BACTERIAL DIMETHYLSULFONIOPROPIONATE DEGRADATION GENES IN THE OLIGOTROPHIC NORTH PACIFIC SUBTROPICAL GYRE <sup>1</sup>

<sup>&</sup>lt;sup>1</sup> Varaljay, V.A., S.M. Gifford, S.T. Wilson, S. Sharma, D.M. Karl, M.A. Moran. 2012. Applied and Environmental Microbiology. 78: 2775-2782. Reprinted here with permission of publisher.

## Abstract

Dimethylsulfoniopropionate (DMSP) is an organic sulfur compound that is rapidly metabolized by marine bacteria either by cleavage to dimethylsulfide (DMS) or demethylation to 3-methiolpropionate. The abundance and diversity of genes encoding bacterial DMS production (*dddP*) and demethylation (*dmdA*) were measured in the North Pacific Subtropical Gyre (NPSG) between May 2008 and February 2009 at Station ALOHA (22° 45'N, 158° 00'W) at two depths: 25 m and the Deep Chlorophyll Maximum (DCM;  $\sim 100$  m). The highest abundance of *dmdA* genes was in May 2008 at 25 m with  $\sim 16.5\%$  of cells harboring a gene in one of the eight subclades surveyed, while the highest abundance of *dddP* genes was in July 2008 at 25 m with ~2% of cells harboring a gene. The *dmdA* gene pool was consistently dominated by homologs from SAR11 subclades, which was supported by findings in metagenomic datasets derived from Station ALOHA. Expression of the SAR11 *dmdA* genes was low, with typical transcript:gene ratios between 1:350 and 1:1,400. Abundance of DMSP genes was statistically different between 25 m and the DCM and correlated with a number of environmental variables including primary production, photosynthetically active radiation, particulate DMSP, and DMS concentrations. At 25 m, *dddP* abundance was positively correlated with pigments that are diagnostic of diatoms; at the DCM, *dmdA* abundance was positively correlated with temperature. Based on gene abundance, we hypothesize that SAR11 bacterioplankton dominate DMSP cycling in the oligotrophic NPSG, with lesser but consistent involvement of other members of the bacterioplankton community.
# Introduction

Dimethylsulfoniopropionate (DMSP) is a common sulfur compound produced by phytoplankton for use as an osmolyte (40, 60). When released into seawater, DMSP is rapidly sequestered and degraded by members of the bacterioplankton community via two major metabolic pathways (28-30). The majority of DMSP (50-90%) is demethylated to 3-methiolpropionate (MMPA), ultimately producing sulfur and carbon intermediates which are incorporated into microbial biomass or further oxidized (29, 43). A competing metabolic pathway results in the production of dimethylsulfide (DMS) from DMSP (16, 21). DMS represents a major source of biogenic sulfur to the atmosphere, where oxidation products form cloud condensation nuclei and ultimately influence radiative backscatter (2, 33, 48).

Recent insights into the molecular mechanisms that drive bacterial DMSP degradation have provided an improved understanding of DMSP cycling at the genomic and transcriptional levels (5, 18, 43, 54, 56, 57). The identification of the DMSP demethylase gene (dmdA), which encodes the first step in the demethylation pathway, has enabled quantification of the gene in marine metagenomic surveys and revealed it to be taxonomically diverse and highly abundant (present in >50% of marine bacterioplankton; ref. 19). To date, dmdA homologs, represented by 5 clades and 14 subclades, are known to be harbored by SAR11, roseobacters, Gammaproteobacteria, and SAR116 member *Candidatus* 'Puniceispirillum marinum' IMCC1322 (19, 20, 62). While there is strong cohesion in dmdA amino acid sequences, there is extensive heterogeneity based on nucleotide sequences (19, 62). In comparison to dmdA, the genes involved in DMS production (dddD, dddL, dddP, dddQ, dddY, and dddW, all of which mediate the same

step of DMSP cleavage) are present in less than 10% of bacteria based on marine metagenomic surveys (9, 10, 19, 54-57). The most abundant *ddd* genes in bacterial taxa are *dddP* and *dddQ*, occurring in genomes of some roseobacters (19, 54, 55) and SAR116.

The objective of this study was to measure the distributions of genes diagnostic of DMSP degradation in the North Pacific Subtropical Gyre (NPSG), the world's largest biome with a surface area of  $1 \times 10^7 \text{ km}^2$  (25). Station (Stn) ALOHA is located in the NPSG at 22° 45'N, 158° 00'W and represents the sampling site of the Hawaii Ocean Time-series (HOT; refs. 23, 25) where a suite of biogeochemical and physical parameters are measured on a near-monthly basis to characterize the long-term biogeochemical cycling in this oligotrophic oceanic ecosystem. Here, we measured *dmdA* and *dddP* abundance and expression at two depths, at 25 m in the nutrient-depleted upper euphotic zone and at the persistent deep chlorophyll maximum (DCM) typically located at ~100 m in the lower euphotic zone (24). Over a 10 month period, genetic and chemical analyses were conducted to identify the bacterial taxa and environmental factors that potentially influence DMSP fate in this ecosystem, including DMS and particulate DMSP (DMSPp) concentrations. Our results indicate that the targeted DMSP-degrading genes are abundant in surface waters of the NPSG, particularly those from the SAR11 clade, and that they show greater depth variability than temporal variability.

## Materials and methods

### **Biogeochemical parameters**

Samples were collected at Stn ALOHA between May 2008 and February 2009 during HOT cruises 201-209. The core physical, chemical and biological measurements were conducted as part of the HOT program as previously described (25, 26) and are available in the Hawaii Ocean Time-series Data Organization and Graphical System (HOT-DOGS; http://hahana.soest.hawaii.edu/hot/ hot-dogs/interface.html). The mixed layer depth was calculated using the 0.125 potential density criterion (38). The average pH and salinity was 8.05 and 35.19, respectively, over the sampling period. Seawater concentrations of DMS and DMSP were quantified on-board using a cryogenic purgeand-trap technique followed by gas chromatography, as previously described (59). In brief, filtered seawater samples were sparged with helium and trapped in a sample loop maintained in liquid nitrogen. The sample loop was subsequently heated and DMS quantified using a gas chromatograph (Agilent 7890) equipped with a flame photometric detector (FPD) and a Chromasil 330 chromatography column. To measure particulate DMSP (DMSPp), a proxy for DMSP within phytoplankton, seawater samples were gently filtered through glass fiber filters, and the filters were exposed to NaOH (1 M) in gas-tight vials to hydrolyze DMSP to DMS, which was measured as described above.

# Nucleic acid collection and extraction

Samples for DNA and RNA analysis were collected from 25 m and the DCM, which ranged from 110 to 140 m during the sampling period (Table 3.1 in Appendix B). Samples were obtained in triplicate from the same conductivity-temperature-depth (CTD) cast as the DMSP and DMS measurements and filtered through 25 mm diameter 0.2 µm

pore-size Durapore filters. Approximately 1.8 L of seawater were filtered for each DNA replicate and 1.4 -1.9 L for RNA replicates. For RNA samples, the filtration time averaged 15 min and was never >20 min. Immediately following filtration, filters were placed into 2 ml cryogenic vials and 250 µl lysis buffer (20 mM Tris HCl, 2 mM EDTA, 1.2% Triton X and 20 mg/ml lysozyme) was added to DNA filters and 250 µl RLT buffer (Qiagen, Valencia, CA) to RNA filters. All tubes, particularly RNA samples, were then immediately flash frozen in liquid nitrogen and subsequently stored at -80 °C. DNA extractions were carried out with the DNeasy Blood and Tissue kit (Qiagen) following the pre-treatment protocol for Gram<sup>+</sup> cells. RNA extractions were carried out using the RNeasy Mini kit (Qiagen). Immediately prior to extraction, an additional 600 µl of RLT buffer and 0.5 ml silica carbide beads (MO BIO laboratories Inc., Carlsbad, CA) were added to the sample tube, which was placed on a bead beating vortex adapter for 5 min. The lysate was transferred to a new tube and the rest of the extraction protocol conducted as described in the RNeasy manual. Following extraction, RNA samples were DNase digested using the Turbo DNA-free kit (Ambion, Austin, TX) with double the enzyme volume. DNA and RNA samples used in qPCR were quantified by Quant-iT PicoGreen and RiboGreen (Invitrogen Ltd., Carlsbad, CA) kits respectively using a TBS-380 fluorometer (Promega, Sunnyvale, CA) or for some RNA samples by a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc, Wilmington, DE). Average DNA concentrations ranged from  $0.2 \pm 0.15$  micrograms per L at 25 m and  $0.17 \pm 0.12$ micrograms per L at the DCM. Average RNA concentrations ranged from  $0.18 \pm 0.17$ micrograms per L at 25 m and  $0.05 \pm 0.04$  micrograms per L at the DCM. The lower

concentration samples (< 5 ng/ $\mu$ l) may have greater measurement error, particularly for NanoDrop quantification.

# Quantitative and RT-quantitative PCR.

The primer set sequences for *dmdA* subclades A/1, A/2, B/3, B/4, C/2, D/1, D/3, and E/2 primer and the corresponding annealing temperatures are described in Varaljay et al. (62), and the *dddP* primer set targeting Group 1 is described in Levine et al. (32). The 16S rRNA BACT1369F and PROK1492R primers were from Suzuki et al. (52). As previously described (32), all quantitative PCR reactions were run in duplicate in 25.0 µl volumes with iQ<sup>™</sup> SYBR Green Supermix (Bio-Rad, Hercules, CA) and 300 nM final primer concentrations. qPCR product size and specificity were verified by agarose gel electrophoresis. For reverse transcription (RT)-qPCR assays, 1X iScript One-Step RT-PCR kit with SYBR Green was used with specific priming cDNA reverse transcription, as specific priming is more sensitive than random hexamer priming. Additionally, for RT-qPCR, a 1.0 ng/µl final concentration of T4 gene 32 protein (Roche Applied Science, Indianapolis, IN) was added to decrease PCR inhibition (8, 32). For DNA and RNA samples,  $3.0 \,\mu$ l and  $5.0 \,\mu$ l template were added to each reaction, respectively. All reactions were run on an iCycler iQ (Bio-Rad, Hercules, CA) with the following cycling conditions: initial denaturation at 95 °C for 5 min, 40 cycles of 95 °C for 30 sec, specified annealing temperature (32, 62) for 30 sec, extension at 72 °C for 30 sec, and then a final denaturation and annealing for 1 min each and a melt curve following each run. RTqPCR cycling conditions were the same, except a 10 min 50 °C reverse transcription step was included prior to the initial denaturation. The limit of detection was determined by: a) amplification above the lowest concentration of the standard curve (ranging from 5-

300 copies per reaction across primer sets), b) specificity of melt curves, and c) minimal contamination ( $\geq$  3 cycle difference) in the no-template or negative reverse-transcriptase controls compared to the sample reactions. To normalize for any differences in extraction efficiency, all gene data were analyzed as a percent of cells harboring the gene based on 16S rRNA qPCR (assuming 1.4 16S rRNA genes per cell as calculated from Stn ALOHA metagenomic 16S rRNA: *recA* estimates; see below).

All qPCR standards were constructed from TOPO TA (Invitrogen, Carlsbad, CA) plasmid clones with PCR gene product inserts. Product inserts and specificity of primer sets were previously verified for *dmdA* using 454-sequencing (as described in ref. 62) and for *dddP* using Sanger sequencing (as described in ref. 32). Five product inserts for 16S rRNA were sequenced from DNA obtained in this study and one of these was used as a qPCR standard. 10-fold serially diluted standards were run on every DNA and RNA plate with an average  $r^2$  of 0.997 and efficiency of ~94% across all standard curves.

### Multi-Dimensional Scaling and statistical analyses

The gene abundance and biogeochemical datasets were compared using Multi-Dimensional Scaling (MDS) in R (53) using the vegan package and Bray-Curtis similarities (39). All gene data were normalized on a scale from 0-1 for the MDS analysis. Subclade B/4 was removed from analyses due to low abundance and nonspecific amplification. Three axes were chosen based on a significant decrease in stress; however only the first and second axes, which contributed the most to the gene distribution variability (81% and 12%, respectively) were finally considered. The MDS output consisted of two matrices of similarity scores: a) sample similarity scores based on the gene abundance patterns (n=15), and b) gene similarity scores based on the sample patterns (n=8). Both sets of scores were plotted on the MDS axes. Since these data did not appear normally distributed, non-parametric Spearman's rank correlations were performed between MDS axes and available environmental data, including depth, month, DMSPp, DMS, chlorophyll a (Chl a), PAR, dissolved organic carbon (DOC), silicate, nitrate + nitrite, temperature, salinity, primary production, fucoxanthin, 19 butanoyloxyfucoxanthin, 19` hexanoyloxyfucoxathin, picoeukaryote abundance, heterotrophic bacterial abundance, and cyanobacterial Synechococcus and *Prochlorococcus* abundance. Statistically significant correlations were determined using a Student's t-test on the Spearman's correlation coefficient ( $\rho$ ) at P < 0.05. Differences in gene abundance and environmental parameter means between samples collected from 25 m and the DCM were verified using a two-sample t-test at P < 0.05. To negate the effects of autocorrelation of environmental data with depth, the samples were divided by depth and intra-depth patterns were identified based on rank correlations carried out between gene abundance and environmental variables using a significance level of P < P0.05.

#### *dmdA* metagenomic and metatranscriptomic analyses

Metagenomic and metatranscriptomic datasets from Stn ALOHA obtained on cruises HOT140 (12), HOT154 (12), HOT175 (41), HOT179 (14, 47), and HOT186 (50) and available in CAMERA (http://camera.calit2.net/) and NCBI's Short Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) were mined for sequences representing *dmdA*, *dddP*, *dddW*, *dddQ*, *dddY*, *dddD*, and *dddL*, and for comparison, proteorhodopsin. Query sequences consisted of full length protein sequences from cultured organisms (see accession numbers in Table 3.2 in Appendix B). All functional gene BLASTs were

carried out using tBLASTn with an E-value cut-off of < 10-4. To increase annotation confidence, all hits were compared to NCBI's RefSeq database in a BLASTx analysis using a bit score cut-off of  $\geq$  40 and manually verified for correct target function. For *dmdA* hits, an additional BLASTx analysis was done to identify clade affiliation (62), also using a bit score cut-off of  $\geq$  40, against a > 3,000 member in-house DmdA (2,440 sequences) and GcvT (623 outgroup sequences) database consisting of cultured and metagenomic sequences from the GOS dataset (45, 63). This was found to be the best approach for assigning short sequences (62), since a test using pplacer (37) with randomly trimmed 100 bp known *dmdA* or *gcvT* sequences miss-assigned 20% of the sequences also limited the ability to accurately determine the percent of sequences with matches to our qPCR primers.

To calculate the percent of cells harboring DMSP-degrading genes, the ratio of 16S rRNA genes per *recA* was determined using E. coli K-12 16S rRNA and *recA* genes (4, 19) as queries in BLASTn and tBLASTn, respectively, with an E-value cutoff of  $<10^{-4}$ . Only those sequences with the correct target annotation and a reciprocal best hit with a bit score  $\geq 100$  for 16S rRNA genes (BLASTn) and a bit score  $\geq 40$  for RecA (BLASTx) were retained. In order to normalize for effects of gene size on the number of hits retrieved, 16S rRNA (1,542 bp) and *recA* (1,059 bp) genes were length-normalized according to Biers et al. (4), which resulted in an average ratio of 1.4:1 for 16S rRNA genes copy numbers as a percent of cells, assuming a single copy of *recA* and 1.4 copies of 16S rRNA genes per cell. While multiple DMSP gene copies per cell could lead to inflated

per cell estimates, SAR11 *Candidatus* 'Pelagibacter ubique' HTCC7211 strain is the only marine isolate known to carry two *dmdA* gene copies (62).

## Results

# HOT physical and biogeochemical data

During the sampling period from May 2008 to February 2009, the depth of the surface mixed layer ranged from 22 to 114 m. With the exception of May 2008, seawater samples collected from 25 m were always located within the mixed layer (Fig. 3.1). The seawater temperatures at 25 m were 1-4 °C higher than those in the DCM (Fig. 3.7A in Appendix B). The daily integrated flux of photosynthetically available radiation (PAR) at 25 m represented approximately 11% of surface irradiance, while the DCM was always situated below the 1% light level (Fig. 3.7B in Appendix B). Primary production was ~13-fold higher at 25 m than at the DCM (Fig. 3.7C in Appendix B). Nitrate + nitrite concentrations were higher at the DCM (11 - 1075 nmol L<sup>-1</sup>) than at 25 m (2 - 6 nmol L<sup>-1</sup>) (Fig. 3.7D in Appendix B).

With respect to DMS(P) chemistry, DMS and DMSPp concentrations were highest in the upper 75 m of the water column and subsequently declined with depth to near-detection limits at 150 m (Fig 3.1). For the discrete depths sampled for gene analysis, DMS and DMSPp concentrations were 2- to 6-fold higher at 25 m compared to the DCM (Fig. 3.2A and B). Between May 2008 and February 2009, DMS concentrations at 25 m were highest from May to October (2.3 - 2.8 nmol L<sup>-1</sup>) and lowest from November to February (1.9 - 2.0 nmol L<sup>-1</sup>) (Fig. 3.2A). At the DCM, DMS concentrations were lower and less variable (0.3 - 0.9 nmol L<sup>-1</sup>) (Fig. 3.2A). The

(Fig. 3.2B). Chl *a* concentrations ranged from ~1.5-fold to 3-fold higher at the DCM compared to 25 m (Fig 3.2C). The DMSPp:Chl *a* ratio (an indicator of the phytoplankton DMSP content per chlorophyll content) was up to 25-fold higher at 25 m (Fig 3.2D) and the highest DMSPp:Chl *a* ratios occurred in May and August 2008 (Fig 3.2D).

## dmdA and dddP abundances

Eight *dmdA* subclades were selected for analysis with PCR primers, targeting ~50% of known *dmdA* sequences (62), and copy numbers were normalized to 16S rRNA qPCR (based on a calculation of 1.4 16S rRNA genes per cell; see Methods). A higher proportion of bacterioplankton cells contained a *dmdA* gene from one of the 8 targeted subclades at 25 m ( $13\% \pm 2\%$ ) than at the DCM ( $6.5\% \pm 1\%$ ) throughout the sampling period (Fig. 3.3). The maximum frequency of *dmdA*-containing cells was 16.5%, occurring in May 2008 at 25 m. Like *dmdA*, a higher fraction of *dddP*-containing cells was observed at 25 m than at the DCM, although this gene was present at a consistently lower frequency than *dmdA* at every sampling occasion and depth. The maximum frequency of cells containing the targeted *dddP* subclade was 2.1%, occurring in July 2008 at 25 m (Fig. 3.3).

The SAR11 subclades made up the greatest portion of *dmdA* genes, with subclades D/1 and C/2 responsible for 80% of the total *dmdA* genes measured. Together, these subclades were present in approximately 10% and 5% of cells at 25 m and the DCM, respectively (Fig. 3.3). *dmdA* genes from the Roseobacter subclade A/2 and Gammaproteobacteria subclade E/2 were particularly enriched in cells inhabiting surface waters, with an order of magnitude difference in their abundance relative to the DCM. All three SAR11 *dmdA* subclades, including the less abundant D/3, were most frequent in

the bacterioplankton community at 25 m in May 2008, while the A/1, A/2 and E/2 subclades were most frequent at 25 m in October 2008. SAR116 *dmdA* subclade B/4 sequences were consistently below the detection limit.

# MDS and statistical analyses

An MDS plot was used to explore *dmdA* and *dddP* profiles by depth and sampling occasion (Fig. 3.4). Together, the first two MDS axes represented ~ 93% of the variability in abundance of measured DMSP-related genes. Samples were strongly separated by depth (Fig. 3.4). MDS scores for the individual genes indicated that surface subclades A/2 and E/2 (Roseobacter and Gammaproteobacteria) grouped together on the MDS plot (Fig. 3.4) and were highly correlated ( $\rho = 0.95$ , P < 0.001). Subclade B/3, which represents an unknown taxonomic group, was typically higher in abundance at the DCM (Fig. 3.4), and was negatively correlated with E/2 and D/1 ( $\rho < -0.56$ , P < 0.03) and not correlated with any other *dmdA* subclade or *dddP*.

To explore the factors that might be driving gene patterns, MDS axis scores were analyzed against biogeochemical parameters. MDS axis 1 correlated positively with depth ( $\rho = 0.87$ , P < 0.001) and less strongly but negatively with several other parameters that had highest values at 25 m, including DMSP:Chl *a* ratio, DMSPp concentration, DMS concentration, temperature, PAR, primary production rate, DOC concentration, heterotrophic bacterial abundance, *Prochlorococcus* and *Synechococcus* abundance, and total DMSP gene counts (all  $\rho < -0.58$ , P < 0.05). In contrast, 19` butanoyloxyfucoxanthin, 19` hexanoyloxyfucoxanthin, nitrate + nitrite, Chl *a*, and silicate concentrations were positively correlated with MDS axis 1, being lower at 25 m than at the DCM (all  $\rho > 0.60$ , P < 0.05). MDS axis 2 was negatively correlated with fucoxanthin, a diagnostic pigment for diatoms ( $\rho = -0.51$ , P < 0.05) and with the DMSPp:Chl *a* ratio (i.e., an index of phytoplankton DMSP content per Chl *a* content;  $\rho = -0.52$ , P < 0.05), but this axis accounted for a smaller fraction of the variability (12%; Fig. 3.4).

To eliminate the strong influence of depth-related environmental signals that dominated axis 1 of the MDS plot, data were also analyzed independently for each depth. At 25 m, overall *dmdA* gene abundance (all subclades combined) was not significantly correlated with any variable, although surface subclades A/2 and E/2 were negatively correlated with the diatom pigment fucoxanthin (both  $\rho$ <-0.85, *P* < 0.05) and subclade A/1 was positively correlated with Chl *a* ( $\rho$  = 0.93, *P* < 0.001). *dddP* gene counts were positively correlated with fucoxanthin ( $\rho$  = 0.78, *P* < 0.05). At the DCM, *dmdA* was positively correlated with temperature ( $\rho$ =0.79, *P* < 0.05) while *dddP* was not significantly correlated with any of the environmental parameters measured.

### *dmdA* gene expression

Select samples from three dates (August 2008, October 2008, and January 2009) and both depths (25 m and the DCM) were analyzed for *dmdA* transcript levels (SAR11 D/1 and D/3 subclades only), but most were at or below the detection limit. In the October 2008 samples from 25 m, however, both D/1 and D/3 subclades had measurable expression. Average transcript:gene ratios were 1:350 for D/3 and 1:1,400 for D/1 in samples for which transcription could be accurately quantified.

# Metagenomic dataset analysis

A homology search of *dmdA* sequences against available metagenomic and metatranscriptomic datasets from Stn ALOHA (12, 14, 41, 47, 50) resulted in hits to

Clade A (Roseobacter), Clades C and D (SAR11), and Clade E (Gammaproteobacteria), with the majority of sequences having best matches to the two SAR11 clades (>75%; Fig. 3.5). From estimates of *recA* abundance in the metagenomic datasets and assuming a single copy of *dmdA* and *recA* per cell, the percentage of cells carrying a *dmdA* gene was  $\sim$ 40%, three-fold higher than the qPCR-based average of  $\sim$ 13% for the 8 primer sets combined. The abundance of cells harboring a *dddP* gene in the HOT metagenomes was 4%, while our qPCR estimate averaged 1.4%. The percentage of cells harboring a *dddD* or a *dddQ* was  $\sim$ 3% or 1.6% respectively. Reads with significant homology to *dddW*, *dddY*, and *dddL* sequences were not detected in the Stn ALOHA metagenomes or metatranscriptomes.

### Discussion

The goals of this study were to determine whether bacterioplankton DMSP genes in the NPSG co-vary with physical or chemical parameters that indicate the environmental conditions conducive for DMSP-relevant microbial processes, and whether variations in taxonomic affiliations of genes over time or space could signal shifts in the dominant bacterial taxa mediating DMSP cycling. Previously, both primary production and solar radiation have been hypothesized to influence DMSP and DMS production. The former is thought to track with rates of biosynthesis of DMSP by phytoplankton cells (3, 36). The latter assumes a role for DMSP in cellular scavenging of reactive oxygen species (49, 51, 61), and indeed solar radiation levels have been linked to increased assimilation of DMSP by microorganisms in the surface mixed layer of the NPSG (11). In this study, primary productivity and solar radiation may also have driven the consistently higher abundance of bacterioplankton cells harboring one of the *dmdA* 

subclades or *dddP* at 25 m compared to the DCM (Fig. 3.3), but many other depth-related parameters also showed strong vertical structure at Stn ALOHA (12, 22, 23) including DMS, DMSPp, Chl *a*, dissolved organic carbon concentrations, DMSP:Chl *a* ratios, and temperature. Because of autocorrelation with depth, these could not be individually resolved in this study.

At Stn ALOHA, non DMSP-producing Prochlorococcus are the dominant phytoplankton (6, 7, 65), with DMSP-producing species such as diatoms and prymnesiophytes present in lower abundance (1, 31, 46). The pigment-related correlations emerging for some DMSP-degrading genes at 25 m (Fig. 3.8 in Appendix B), however, might indicate ecological interactions between DMSP-degrading bacteria and one of the DMSP-producing phytoplankton groups, such as the positive correlation between *dddP* gene abundance [known to be harbored by Roseobacter (ref. 54) and SAR116 cells thus far] and fucoxanthin, a pigment diagnostic of diatom cells. The DMSPp:Chl a ratio is considered an indicator of the fractional importance of DMSP in the available organic carbon pool (27, 44) and is expected to track with the abundance of high DMSP-producing phytoplankton species within a given light regime. However, on the two sampling occasions during which the DMSPp:Chl a ratio was 3-fold higher than average (Fig. 3.2; 25 m in May and August 2008), there were no obvious changes in composition or abundance of the DMSP gene pool targeted with the qPCR primer sets (Fig. 3.3).

SAR11 clades (particularly clade D/1) dominated the *dmdA* pool at every depth and station in this study, similar to what was found over 36 ocean surface waters surveyed in the 2007 GOS dataset (19) and consistent with the recognized abundance of

SAR11 in the Stn ALOHA bacterioplankton (13). Two other *dmdA* clades, Roseobacter clade A/2 and Gammaproteobacteria clade E/2, were largely confined to the upper mixed layer, while subclade B/3 (taxonomic affiliation unknown) was primarily a DCM indicator. These patterns of taxonomic affiliations of DMSP-related genes at Stn ALOHA were not correlated to DMSP or DMS pool sizes at either 25 m or the DCM, but might nonetheless signal differences in the dominant pathways or processing rates. For example, SAR11 bacteria rely on reduced sulfur for growth (58) and are capable only of DMSP demethylation, based on the genome sequences available thus far, while some roseobacters are able to both demethylate and cleave DMSP (16, 19, 35). These overall findings are also consistent with a study of the upper 60 m of another oligotrophic ocean gyre, the Sargasso Sea, in which SAR11 *dmdA* genes dominate year-round (32).

The average *dmdA* and *dddP* frequencies obtained by qPCR analysis were systematically lower than those estimated from the HOT179 metagenomic data, but this is attributable to the requirement for highly conserved nucleotide sequences for qPCR primer design and, as a consequence, the fact that the eight qPCR primer sets used here targeted only about half of the currently known *dmdA* sequences from the 2007 GOS metagenome (62). Furthermore, *dmdA* groups not represented in the GOS 2007 may be present at Stn ALOHA, particularly since the DCM habitat was not sampled in the GOS (45). Given the constraints of environmental primer design, the qPCR and metagenomic estimates were in good agreement. Both the HOT179 metagenome and qPCR data similarly found that most *dmdA* genes in NPSG surface waters were associated with SAR11 clades (~75% for the HOT179 and 80% for qPCR). In addition, both methods found an order of magnitude difference in *dmdA* and *dddP* abundance (13:1 for the

HOT179 metagenome and 9:1 for qPCR), which may be a common feature of marine environments (19, 32, 43, 54). Field studies indicate that ~80% of bacteriallymetabolized DMSP is processed through the demethylation pathway and only ~20% is cleaved to DMS (29). Evolutionary pressure for marine bacteria to maintain the ability to demethylate DMSP could explain the consistently high and stable dmdA gene frequencies found in ocean bacteria, including those at Stn ALOHA. Transcript levels in NPSG surface waters, which should better correlate with the conditions under which the gene products are ecologically advantageous, were also quantified, but only the October samples had expression levels above the limit of detection for dmdA subclades D/1 and D/3.

Poor transcript detection for DMSP-related genes may be due in part to the relatively low RNA concentrations recovered from Stn ALOHA (many RNA yields were below 100 ng L<sup>-1</sup>), or detection may appear low if the DNA pool is artificially inflated by detrital material (17, 66). However, there is also evidence that *dmdA* has inherently low expression relative to other bacterial genes at Stn ALOHA, since low transcript abundance was also found in the HOT179 metatranscriptomic data. For example, the expression ratio in the HOT179 sequence libraries [calculated as % representation in the cDNA library/% representation in the DNA library, according to Frias-Lopez et al. (ref. 14), which is a relative expression ratio] was 0.11, whereas the expression ratio calculated in the same way for proteorhodopsin, another high frequency SAR11-dominated gene in the bacterioplankton DNA, was 100-fold higher (Fig. 3.9 in Appendix B). Since a diel survey conducted during the October cruise did not find significant variation in transcript abundance across 8-10 time points over 48 h (Fig. 3.6), low *dmdA* 

expression levels cannot simply be attributed to inopportune sampling times. However, storage of undegraded DMSP by bacterioplankton for use as an osmolyte (42), short mRNA half-lives, or a highly stable or efficient *dmdA* protein product might uncouple instantaneous gene transcription rates from DMSP turnover rates. Low expression levels for *dmdA* have been detected in a number of other marine systems using qPCR and metatranscriptomics (15, 32, 64). Understanding the relationship between gene frequency, transcript abundance, protein levels, and biogeochemical rates is a crucial challenge for future research.

Overall, *dmdA* and *dddP* gene pools targeted by our suite of primers at Stn ALOHA showed greater variation between the surface mixed layer and the DCM than they did within either depth throughout the 10 month study period (Fig. 3.3), in agreement with the strong vertical structure but low seasonality of the NPSG (34). We propose that the high and relatively invariant inventory of bacterial DMSP genes in the NPSG indicates strong evolutionary pressure on bacterioplankton to maintain this capability, and that DMSP degradation is not the purview of specialized bacteria. Based on the composition of the DMSP gene pool, SAR11 bacterioplankton dominate DMSP cycling in the upper ocean of the oligotrophic NPSG throughout the year, with lesser but consistent involvement of members of the Roseobacter and Gammaproteobacteria taxa.

## Acknowledgements

We would like to thank the Captain and crew of the R/V *Kilo Moana* and the Hawaii Ocean Time-series (HOT) for availability of environmental data, and M. Church for additional DNA samples collected during 2008 for qPCR verification. N. Levine, H. Luo, R. Newton and S. Holland provided helpful discussions and advice on statistical

analyses. This work was supported by grants from the Gordon and Betty Moore Foundation, NSF (EnGen award OCE0724017), and the NSF-supported Center for Microbial Oceanography: Research and Education (C-MORE) (EF0424599).

# References

- Andersen, R. A., R. R. Bidigare, M. D. Keller, and M. Latasa. 1996. A comparison of HPLC pigment signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. Deep-Sea Res. Pt. II 43:517-537.
- Andreae, M. O., and P. J. Crutzen. 1997. Atmospheric aerosols: biogeochemical sources and role in atmospheric chemistry. Science 276:1052-1058.
- Bell, T. G., A. J. Poulton, and G. Malin. 2010. Strong linkages between dimethylsulphoniopropionate (DMSP) and phytoplankton community physiology in a large subtropical and tropical Atlantic Ocean data set. Global Biogeochem. Cy. 24:1-12.
- Biers, E. J., S. Sun, and E. C. Howard. 2009. Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. Appl. Environ. Microbiol. 75:2221-2229.
- Bürgmann, H., E. C. Howard, W. Ye, F. Sun, S. Sun, S. Napierala, and M. A. Moran. 2007. Transcriptional response of *Silicibacter pomeroyi* DSS-3 to dimethylsulfoniopropionate (DMSP). Environ. Microbiol. 9:2742-2755.
- Campbell, L., H. A. Nolla, and D. Vaulot. 1994. The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. Limnol. Oceanogr. 39:954-961.

- Campbell, L., and D. Vaulot. 1993. Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA). Deep-Sea Res. Pt. I 40:2043-2060.
- Chandler, D. P., C. A. Wagnon, and H. Bolton, Jr. 1998. Reverse transcriptase (RT) inhibition of PCR at low concentrations of template and its implications for quantitative RT-PCR. Appl. Environ. Microbiol. 64:669-677.
- Curson, A. R. J., R. Rogers, J. D. Todd, C. A. Brearley, and A. W. B. Johnston. 2008. Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine αproteobacteria and *Rhodobacter sphaeroides*. Environ. Microbiol. 10:757-767.
- Curson, A. R. J., M. J. Sullivan, J. D. Todd, and A. W. B. Johnston. 2011.
   DddY, a periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse species of *Proteobacteria*. ISME J. 5:1191-1200.
- Del Valle, D. A., R. P. Kiene, and D. M. Karl. 2012. Effect of visible light on dimethylsulfoniopropionate assimilation and conversion to dimethylsulfide in the North Pacific Subtropical Gyre. Aquat. Microb. Ecol. 66:47-62
- 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.
- Eiler, A., D. H. Hayakawa, M. J. Church, D. M. Karl, and M. S. Rappé. 2009.
   Dynamics of the SAR11 bacterioplankton lineage in relation to environmental

conditions in the oligotrophic North Pacific subtropical gyre. Environ. Microbiol. **11**:2291-2300.

- Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. DeLong. 2008. Microbial community gene expression in ocean surface waters. P. Natl. Acad. Sci. USA. 105:3805-3810.
- Gifford, S. M., S. Sharma, J. M. Rinta-Kanto, and M. A. Moran. 2011.
   Quantitative analysis of a deeply sequenced marine microbial metatranscriptome.
   ISME J. 5:461-472.
- Gonzalez, J. M., R. P. Kiene, and M. A. Moran. 1999. Transformation of sulfur compounds by an abundant lineage of marine bacteria in the α-subclass of the class *Proteobacteria*. Appl. Environ. Microbiol. 65:3810-3819.
- Holm-Hansen, O., W. H. Sutcliffe Jr., and J. Sharp. 1968. Measurement of deoxyribonucleic acid in the ocean and its ecological significance. Limnol. Oceanogr. 13:507-514.
- Howard, E. C., J. R. Henriksen, A. Buchan, C. R. Reisch, H. Bürgmann, R. Welsh, W. Y. Ye, J. M. Gonzalez, K. Mace, S. B. Joye, R. P. Kiene, W. B. Whitman, and M. A. Moran. 2006. Bacterial taxa that limit sulfur flux from the ocean. Science 314:649-652.
- Howard, E. C., S. Sun, E. J. Biers, and M. A. Moran. 2008. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. Environ. Microbiol. 10:2397-2410.
- Howard, E. C., S. Sun, C. R. Reisch, D. A. del Valle, H. Bürgmann, R. P.
   Kiene, and M. A. Moran. 2011. Changes in dimethylsulfoniopropionate

demethylase gene assemblages in response to an induced phytoplankton bloom. Appl. Environ. Microbiol. **77**:524-531.

- Johnston, A. W. B., J. D. Todd, L. Sun, M. N. Nikolaidou-Katsaridou, A. R. J. Curson, and R. Rogers. 2008. Molecular diversity of bacterial production of the climate-changing gas, dimethyl sulphide, a molecule that impinges on local and global symbioses. J. Exp. Bot. 59:1059-1067.
- Karl, D. M. 2007. Microbial oceanography: paradigms, processes and promises. Nat. Rev. Microbiol. 5:759-769.
- Karl, D. M. 1999. A sea of change: biogeochemical variability in the North Pacific Subtropical Gyre. Ecosystems. 2:181-214.
- Karl, D. M., R. R. Bidigare, and R. M. Letelier. 2002. Sustained and aperiodic variability in organic matter production and phototrophic microbial community structure in the North Pacific Subtropical Gyre, p. 222-264, In P. J. Williams, D. N. Thomas, C.S. Reynolds (ed.), Phytoplankton productivity: carbon assimilation in marine and freshwater ecosystems, Blackwell Science Ltd, Ames, IA.
- Karl, D. M., and R. Lukas. 1996. The Hawaii Ocean Time-series (HOT) program: background, rationale and field implementation. Deep-Sea Res. Pt. II 43:129-156.
- 26. Karl, D. M., C. D. Winn, D. V. W. Hebel, and R. Letelier. 1990. Hawaii Ocean Time-series program. Field and laboratory protocols. School of Ocean and Earth Science and Technology, University of Hawaii.

- 27. Kiene, R. P., and L. J. Linn. 2000. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico. Limnol. Oceanogr. 45:849-861.
- Kiene, R. P., L. J. Linn, and J. A. Bruton. 2000. New and important roles for DMSP in marine microbial communities. J. Sea Res. 43:209-224.
- Kiene, R. P., L. J. Linn, J. Gonzalez, M. A. Moran, and J. A. Bruton. 1999. Dimethylsulfoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. Appl. Environ. Microbiol. 65:4549-4558.
- Ledyard, K. M., and J. W. H. Dacey. 1996. Microbial cycling of DMSP and DMS in coastal and oligotrophic seawater. Limnol. Oceanogr. 41:33-40.
- Letelier, R., R. R. Bidigare, D. V. Hebel, M. Ondrusek, C. D. Winn, and D.
   M. Karl. 1993. Temporal variability of phytoplankton community structure based on pigment analysis. Limnol. Oceanogr. 38:1420-1437.
- 32. Levine, N. M., V. A. Varaljay, D. Toole, A., J. W. H. Dacey, S. C. Doney, and M. A. Moran. 2012. Environmental, biochemical, and genetic drivers of DMSP degradation and DMS production in the Sargasso Sea. Environ. Microbiol. doi: 10.1111/j.1462-2920.2012.02700.x.
- Lovelock, J. E., R. J. Maggs, and R. A. Rasmussen. 1972. Atmospheric dimethyl sulphide and the natural sulphur cycle. Nature 237:452-453.
- 34. Malmstrom, R. R., A. Coe, G. C. Kettler, A. C. Martiny, J. Frias-Lopez, E. R. Zinser, and S. W. Chisholm. 2010. Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. ISME J. 4:1252-1264.

- Malmstrom, R. R., R. P. Kiene, M. T. Cottrell, and D. L. Kirchman. 2004. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic Ocean. Appl. Environ. Microbiol. 70:4129-4135.
- 36. Matrai, P., M. Vernet, and P. Wassman. 2007. Relating temporal and spatial patterns of DMSP in the Barents Sea to phytoplankton biomass and productivity.
   J. Marine Syst. 67:83-101.
- 37. Matsen, E. A., R. B. Kodner, and E. V. Armbrust. 2010. Pplacer: linear time maximum-likelihood and bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538.
- Monterey, G. I., and S. Levitus. 1997. Seasonal variability of mixed layer depth for the world ocean. NOAA NESDIS Atlas 14:5.
- 39. Oksanen, J., R. Kindt, P. Legendre, B. O. H. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2009. vegan: community ecology package. R package version 1.15-4. http://CRAN.R-project.org/package=vegan.
- 40. Pichereau, V., J.-A. Pocard, J. Hamelin, C. Blanco, and T. Bernard. 1998. Differential effects of dimethylsulfoniopropionate, dimethylsulfonioacetate, and other s-methylated compounds on the growth of *Sinorhizobium meliloti* at low and high osmolarities. Appl. Environ. Microbiol. 64:1420-1429.
- Poretsky, R. S., I. Hewson, S. Sun, A. E. Allen, J. P. Zehr, and M. A. Moran.
   2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific Subtropical Gyre. Environ. Microbiol. 11:1358-1375.

- 42. Reisch, C. R., M. A. Moran, and W. B. Whitman. 2008.
  Dimethylsulfoniopropionate-dependent demethylase (DmdA) from *Pelagibacter ubique* and *Silicibacter pomeroyi*. J. Bacteriol. 190:8018-8024.
- 43. Reisch, C. R., M. J. Stoudemayer, V. A. Varaljay, I. J. Amster, M. A. Moran, and W. B. Whitman. 2011. Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. Nature. 473:208-211.
- 44. Rinta-Kanto, J. M., S. Sun, S. Sharma, R. P. Kiene, and M. A. Moran. 2011.
   Bacterial community transcription patterns during a marine phytoplankton bloom.
   Environ. Microbiol. 14:228–239.
- Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, et al. 2007. The *Sorcerer II* global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. 5:e77.
- 46. Scharek, R., M. Latasa, D. M. Karl, and R. R. Bidigare. 1999. Temporal variations in diatom abundance and downward vertical flux in the oligotrophic North Pacific gyre. Deep-Sea Res. Pt. I 46:1051-1075.
- 47. Shi, Y., G. W. Tyson, and E. F. DeLong. 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. Nature **459**:266-269.
- 48. Simo, R. 2001. Production of atmospheric sulfur by oceanic plankton:
  biogeochemical, ecological and evolutionary links. Trends Ecol. Evol. 16:287-294.

- Slezak, D., R. Kiene, D. Toole, R. Simó, and D. Kieber. 2007. Effects of solar radiation on the fate of dissolved DMSP and conversion to DMS in seawater. Aquat. Sci. 69:377-393.
- 50. **Stewart, F. J., E. A. Ottesen, and E. F. DeLong**. 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. ISME J. **4**:896-907.
- 51. Sunda, W., D. J. Kieber, R. P. Kiene, and S. Huntsman. 2002. An antioxidant function for DMSP and DMS in marine algae. Nature **418**:317-320.
- Suzuki, M. T., L. T. Taylor, and E. F. DeLong. 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Appl. Environ. Microbiol. 66:4605-4614.
- 53. R Development Core Team. 2011. R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna Austria. ISBN 3-900051-07-0. http://www.R-project.org.
- 54. Todd, J. D., A. J. Curson, C. L. Dupont, P. Nicholson, and A. W. B. Johnston.
  2009. The *dddP* gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. Environ. Microbiol. 11:1624-1625.
- 55. Todd, J. D., A. R. J. Curson, M. Kirkwood, M. J. Sullivan, R. T. Green, and A. W. B. Johnston. 2011. DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. Environ. Microbiol. 13:427-438.

- 56. Todd, J. D., M. Kirkwood, S. Newton-Payne, and A. W. B. Johnston. 2012. DddW, a third DMSP lyase in a model Roseobacter marine bacterium, *Ruegeria pomeroyi* DSS-3. ISME J. 6:223-226.
- 57. Todd, J. D., R. Rogers, Y. G. Li, M. Wexler, P. L. Bond, L. Sun, A. R. J. Curson, G. Malin, M. Steinke, and A. W. B. Johnston. 2007. Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. Science 315:666-669.
- 58. Tripp, H. J., J. B. Kitner, M. S. Schwalbach, J. W. H. Dacey, L. J. Wilhelm, and S. J. Giovannoni. 2008. SAR11 marine bacteria require exogenous reduced sulphur for growth. Nature 452:741-744.
- 59. Turner, S. M., G. Malin, L. E. Bågander, and C. Leck. 1990. Interlaboratory calibration and sample analysis of dimethyl sulphide in water. Mar. Chem. 29:47-62.
- 60. Vairavamurthy, A., M. O. Andreae, and R. L. Iverson. 1985. Biosynthesis of dimethylsulfide and dimethylpropiothetin by *Hymenomonas carterae* in relation to sulfur source and salinity variations. Limnol. Oceanogr. **30**:59-70.
- 61. **Vallina, S. M., and R. Simo**. 2007. Strong relationship between DMS and the solar radiation dose over the global surface ocean. Science **315**:506-508.
- 62. Varaljay, V. A., E. C. Howard, S. Sun, and M. A. Moran. 2010. Deep sequencing of a dimethylsulfoniopropionate-degrading gene (*dmdA*) by using PCR primer pairs designed on the basis of marine metagenomic data. Appl. Environ. Microbiol. **76**:609-617.

- 63. Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, et al.
  2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science
  304:66-74.
- 64. Vila-Costa, M., J. M. Rinta-Kanto, S. Sun, S. Sharma, R. Poretsky, and M.
   A. Moran. 2010. Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. ISME J. 4:1410-1420.
- 65. Vila-Costa, M., R. Simo, H. Harada, J. M. Gasol, D. Slezak, and R. P. Kiene.
  2006. Dimethylsulfoniopropionate uptake by marine phytoplankton. Science
  314:652-654.
- Winn, C. D., and D. M. Karl. 1986. Diel nucleic acid synthesis and particulate
   DNA concentrations: conflicts with division rate estimates by DNA accumulation.
   Limnol. Oceanogr. 31:637-645.

Figure 3.1. Representative depth profile from August 2008 (HOT204) of DMSP-related biogeochemical data at Stn ALOHA. Peaks in DMSPp concentration at 25 m and Chl *a* concentration at the DCM are indicated by arrows. The mixed layer depth (MLD) is shown as a dotted line.



Figure 3.2. DMSP-related biogeochemical properties at Stn ALOHA for 25 m (open symbols) and the DCM (closed symbols). Data were collected on HOT cruises 201-209.



Figure 3.3. Abundance of *dmdA* and *dddP* in Stn ALOHA bacterioplankton cells at two depths, 25 m (top bar graph) and DCM (110-140 m) (bottom bar graph). *dmdA* abundances are shown in the multicolored bars, with the subclade color codes and phylogenetic relationships (20, 62) indicated at the top of the figure. *dddP* abundances are shown in the light blue bars. The relative gene abundances are calculated using the ratio of *dmdA* or *dddP* to 16S rRNA gene copies, assuming 1.4 copies per cell as calculated from HOT metagenomic datasets (see Methods).



Figure 3.4. Multi-Dimensional Scaling (MDS) plot of *dmdA* and *dddP* gene abundance using R (vegan package and Bray-Curtis similarities). Triangles represent samples from 25 m and squares represent samples from the DCM, with the sample month indicated within each symbol. The variability explained by Axis 1 and Axis 2 is 81% and 12%, respectively. Corresponding MDS scores for each *dmdA* subclade and *dddP* are also plotted.


Figure 3.5. Relative abundance of *dmdA* clades in two metagenomic (top) and three metatranscriptomic (bottom) datasets collected on multiple HOT cruises at Stn ALOHA (12, 14, 41, 47, 50). The five clades (A, B, C, D, and E) plus U (which designates unclassified *dmdA* sequences) are shown as percentage of total *dmdA* hits. The number of hits to each clade is shown in parentheses.



Figure 3.6. Gene and transcript copies at Stn ALOHA over a diel cycle for October 25 m (HOT205) for SAR11 *dmdA* subclades D/1 (top) and D/3 (bottom). Vertical bars are standard deviations for duplicate or triplicate biological samples. Dotted lines are the detection limits. RNA samples were not collected at the 2:38 AM or 3:12 PM samples and expression could not be quantified for D/1 at the 2:51 AM sample or for D/3 at the 12:52, 5:42 or 7:12 PM samples.



## CHAPTER 4

# IN SITU TIME-SERIES MEASUREMENTS OF BACTERIAL DIMETHYLSULFONIOPROPIONATE GENES FROM A MONTEREY BAY COASTAL MOORING<sup>1</sup>

<sup>&</sup>lt;sup>1</sup> Varaljay, V.A., J. Robidart, C.M. Preston, J.P. Ryan, R. Marin III, R.P. Kiene, J.P. Zehr. C.A. Scholin, and M.A. Moran. To be submitted to Environmental Microbiology.

## Abstract

Dimethylsulfoniopropionate (DMSP) is a phytoplankton-derived organic sulfur compound that once released into seawater can be degraded by bacteria in the cleavage pathway (to produce climatically-relevant dimethylsulfide which participates in cloud formation) or in the demethylation pathway (to produce sulfur intermediates for amino acid synthesis or energy generation). In Monterey Bay, CA, we took advantage of the remote analysis capabilities of the Environmental Sample Processor (ESP) to measure bacterial DMSP genes at near-daily resolution. qPCR primers targeted genes from SAR11 cells (*dmdA*, mediating demethylation) and Roseobacter cells (*dmdA* and also *dddP*, mediating cleavage) over the course of several DMSP-producing dinoflagellate dominated events from September 28 to October 28, 2010. Gene abundances were measured *in situ* and averaged 9.0  $\times 10^7$  copies L<sup>-1</sup> for SAR11 *dmdA*, 3.4  $\times 10^7$  for Roseobacter dmdA, and 3.1 x 10<sup>7</sup> for Roseobacter dddP. Transcript abundances were measured post-deployment on samples collected by the ESP and archived on the instrument until retrieval. Expression ratios (transcript copies: gene copies) were significantly correlated with particulate DMSP:Chl a ratios, a relationship suggesting that availability of DMSP as a proportion of phytoplankton-derived labile organic matter is a potential regulatory signal. Roseobacter *dmdA* genes were transcribed at a gene expression level ~3.5-fold higher than SAR11 dmdA, but the SAR11 genes were ~3-fold more abundant. SAR11 *dmdA* gene expression ratios were significantly higher in the free-living size fraction compared to the particle-associated fraction, indicating a lesser role in DMSP cycling for this taxon in enriched seawater microniches. Differences in expression patterns between Roseobacter *dmdA* and *dddP* genes, each of which mediates

the first committed step of the two competing DMSP degradation pathways, may provide a bioassay for the relative importance of cleavage versus demethylation in marine systems.

## Introduction

Dimethylsulfoniopropionate (DMSP) is a sulfur-containing algal osmolyte, antioxidant, and predator deterrent (27, 40, 56, 64, 70, 71). A significant amount of DMSP is released from phytoplankton into seawater dissolved organic carbon and sulfur pools, especially during blooms dominated by coccolithophores or dinoflagellates (5, 15, 53, 63), and is rapidly consumed by marine bacteria on the order of hours to days (22, 73). When bacteria degrade DMSP by the "cleavage pathway", one of the degradation products, dimethylsulfide (DMS), is degassed from ocean surface waters to the atmosphere where it participates in the formation of cloud condensation nuclei (2, 7). Alternatively, bacteria can use the "demethylation pathway" to produce less volatile sulfur compounds, including sulfur-containing amino acids, which are important in the ocean food web (14, 25).

Studies have shown that bacteria consuming DMSP from the dissolved pool route more through the demethylation pathway (50-90%) than through the cleavage pathway (23, 24). While environmental conditions such as availability of reduced carbon and sulfur compounds and solar radiation levels have been proposed to play a role in determining the routing between DMSP degradation pathways (9, 24, 30), it is not yet clearly understood what regulates community switches between these pathways *in situ*. Understanding the environmental cues that favor the expression of a particular DMSP

degradation pathway, especially during DMSP-producing phytoplankton blooms, is critical to predicting DMS flux.

Two abundant groups of bacteria known to degrade DMSP (14, 32), the Roseobacter and SAR11 clades in the Alphaproteobacteria, typically account for ~10% and ~35% of the bacterioplankton in surface waters respectively (4, 36). Members of both groups harbor the DMSP demethylase gene, *dmdA*, which encodes the first step in the demethylation pathway (18, 19). Roseobacter clade members also possess the genes for the competing DMS-producing pathway (*dddP*, *dddQ*, *dddW*, *dddD*, and *dddL*; refs. 8, 59-62).

SAR11 and roseobacters are common taxa in Monterey Bay, CA (46, 57), a coastal upwelling system where seasonal dinoflagellate blooms are prominent and DMSP and DMS cycling is highly variable (29, 39, 49). We used an autonomous remote-sampling instrument, the Environmental Sample Processor (ESP; refs. 42, 43, 52), to determine the abundance and expression of DMSP degradation genes in dominant members of these bacterial lineages in Monterey Bay. *dmdA* and *dddP* genes from relatives of Roseobacter clade strain HTCC2255 and a *dmdA* gene from a SAR11 clade were tracked with near-daily resolution using the autonomous sampling capabilities of the ESP over a month-long period that captured at least two DMSP-producing blooms. We asked whether abundance and expression patterns differed between the Roseobacter and SAR11 groups under changing DMSP and chlorophyll regimes.

#### **Experimental Procedures**

#### Metagenomic-based Primer Design

Marine metagenomic sequence data from Monterey Bay (CAMERA accession CAM\_PROJ\_MontereyBay; 3 samples collected from October 2000, April 2001, and May 2001; http://camera.calit2.net/#; ref. 46) were mined for *dmdA* and *dddP* sequences by tBLASTn queries using full-length DmdA and DddP protein sequences with an E value cutoff of  $10^{-4}$ . Hits were manually verified for the target function using BLASTx and a bit score cut-off of  $\geq$ 40 against NCBI RefSeq for *dddP* or to a >3,000 member inhouse database for *dmdA* (previously determined to accurately distinguish *dmdA* sequences from paralogs, ref. 67). Metagenomic *dmdA* sequences with homology to genes from Roseobacter clade strain HTCC2255 and SAR11 clade D/1 (67) made up ~44% of *dmdA* hits. Metagenomic *dddP* sequences with homology to HTCC2255 sequences made up ~50% of the *dddP* genes detected. These metagenomic sequences were subsequently used in the design of three primer sets: HTCC2255 *dmdA*, HTCC2255 *dddP*, and SAR11 D/1 *dmdA*.

Metagenomic sequences along with PCR products from each primer set amplified from composite Monterey Bay DNA samples collected in 2006 and 2007 were used in the design of corresponding 5' nuclease probes for each primer set (Table 4.1). An *in silico* specificity check showed no non-specific matches of any primer/probe set with non-target *dmdA* or *dddP* genes or with paralogs from the NCBI nucleotide (nt) database, even with a relatively high allowance of 12 mismatches per primer or probe sequence. A previously designed SAR11 16S 5' nuclease probe assay [from Suzuki et. al. (58)] was also used as described in a previous ESP qPCR study (42) (Table 4.1).

## Pre-deployment testing

Annealing temperature gradients and primer concentration matrices were used to determine optimal qPCR assay conditions (Table 4.1). All qPCR assays consisted of 1X Accuprime Supermix I (Invitrogen, Carlsbad, CA), 300 nM hydrolysis probe labeled with FAM/BHQ-1 and 2.5 mM magnesium chloride final concentrations in 30 µl volumes. Reagents were loaded into the ESP, and the qPCR assays were carried out as described previously (42, 48). Cycling conditions were as follows: 95°C for 2.0 min, 42 cycles of 95°C for 15 s and the specified annealing temperature for 1.0 min (see Table 4.1). Prior to deployment, standard curves were run for the primer sets in triplicate 10-fold serial dilutions from  $10^2$  to  $10^5$  gene copies per reaction, except HTCC2255 *dmdA* primers which only had a two point standard curve ( $10^3$  and  $10^4$  gene copies per reaction). Predeployment standard curve efficiencies were 87-98%, and linear regression  $r^2$  values were >0.99 for all DMSP genes and >0.97 for the SAR11 16S primer set. Standards for all qPCR assays were linearized clones with PCR product inserts. Cross-reactivity of standards at  $10^7$ ,  $10^6$ , and  $10^5$  gene copies per reaction between primer sets was minimal, with cross-reactivity < 0.01% for any primer-standard pair.

DNA extraction efficiency of the ESP microfluidic block (MFB) was shown to be comparable with a parallel bench-top extraction of calf thymus DNA using a Qiagen DNeasy kit (Valencia, CA). qPCR tests using surface seawater from the Coastal Data Information Program (CDIP) Station 156 in Monterey Bay and from Monterey Bay Wharf verified that quantification by all primer sets was comparable between the ESP module and a bench-top assay. A more detailed description of ESP procedures can be found elsewhere (42, 43).

## ESP deployment

The ESP was deployed near Station M0 (36.835 N, 121.901 W) at a depth of 8.1 m ( $\pm$  0.7). The ESP was fitted with an SBE 16plus CTD (Sea-Bird, Bellevue, WA) with a Turner Cyclops 7 fluorometer (Turner Designs, Sunnyvale, CA) and a Cstar transmissometer (WET Labs, Philomath, OR) for chlorophyll (Chl *a*), depth, temperature, and beam transmission measurements.

From September 28-October 28, 2010, 15 discrete samples were collected for in situ DNA extraction and qPCR, and 19 samples were collected for filter archiving (Table 4.3 in Appendix C). For *in situ* extraction and real-time qPCR in the ESP, 1.0 L of seawater was collected onto 0.2 µm pore size 25 mm diameter filters and extracted using a modified Qiagen (Valencia, CA) DNeasy extraction method as previously described (42). Primer sets were run in single reactions for each *in situ* time-point with 6.0  $\mu$ l of the extracted DNA template. Negative (no template) controls were run once during the deployment and showed no amplification. Negative control lysates (to check for residual contamination from the DNA extraction column) were run before, after, and at 3 other times during the deployment, and the amplification signal was always less than 5% of the signal for environmental samples. An internal positive control reaction (template included in the primer/probe reagent) was run 14 times to assess the technical variability of the ESP qPCR module and any PCR inhibition. The internal positive control cycle threshold (Ct) value was consistent and did not indicate the presence of inhibition during the deployment [average Ct value =  $29.01 (\pm 0.35)$ ].

Approximately 1-2 hours following the seawater filtration for *in situ* DNA extraction (Table 4.3 in Appendix C), a second 1.0 L seawater sample was filtered in-line

through 5.0  $\mu$ m and 0.2  $\mu$ m pore size 25 mm diameter filters and preserved with two 40 minute sequential incubations of 2.0 ml of RNAlater (Ambion, Austin, TX). Following the deployment, the archived filters were removed from the ESP, flash frozen in liquid nitrogen and stored at -80 °C.

Because it was not possible to run qPCR standards on the ESP during the deployment, standard curves for each primer set were also run post-deployment with the same reagents. Gene quantification was based on the average of pre- and post-deployment standard curves, except HTCC2255 *dmdA* which used the two point pre-deployment standard curve.

## Niskin-based sample collection

During the ESP deployment, ship casts using an SBE 19plus SEACAT CTD (Sea-Bird) with 5.0 L Niskin bottles were used to collect water at an average depth of 9.2 m ( $\pm$  0.7) within 0.5 mi of the ESP for supplementary molecular and chemical measurements (Table 4.3 in Appendix C). Water was returned to the lab usually within 2 h of collection and filtered for nucleic acids by two methods: one for DNA extraction only and one for simultaneous DNA and RNA extraction. For DNA-only filters, triplicate 200-500 ml volumes of seawater were filtered by vacuum filtration, and filters were stored at -80 °C. For DNA and RNA filters, triplicate 1.0 L volumes of seawater were filtered (<30 min filtration time) by peristaltic pump through in-line 5.0  $\mu$ M and 0.2  $\mu$ M pore size filters, except for the October 21 samples for which 0.7 L volumes were filtered by vacuum filtration. All filters were immediately preserved with 100  $\mu$ l of RNAlater (Ambion), flash frozen in liquid nitrogen, and stored at -80 °C.

All biochemical measurements were carried out in triplicate with subsamples from the same Niskin bottle. Chl a was measured from 200 ml of seawater filtered onto Whatman GF/F filters, extracted in 5 ml 90% acetone at -20 °C and quantified by fluorometry (39). Samples for total DMSP (DMSPt) analysis were collected as whole seawater and preserved with HCl (1.5% final concentrations). Samples for dissolved DMSP (DMSPd) analysis were collected by small volume gravity drip filtration through a GF/F filter and then vacuum filtered through a 0.2  $\mu$ m nylon filter and preserved in H<sub>2</sub>SO<sub>4</sub> (1% final concentrations) (26, 54). Particulate DMSP (DMSPp) concentrations were calculated as the difference between DMSPt and DMSPd concentrations and represent the amount of DMSP in phytoplankton cells. DMSPd consumption rates were measured over 4 h or 6 h time courses according to the method of Kiene and Gerard (21). Briefly, whole seawater was incubated in triplicate with or without addition of glycine betaine (+GBT) to inhibit DMSP uptake, and the difference in accumulation rate of DMSPd between the +GBT and control incubations was calculated. The turnover rate constant for the DMSPd pool, k  $(d^{-1})$ , was calculated as the DMSP consumption rate (nM d<sup>-1</sup>) divided by DMSPd concentration (nM). All DMSP measurements were made by cleaving DMSP into DMS with strong alkali and quantifying DMS by gas chromatography.

Slides for phytoplankton taxonomic enumeration were made by filtering 10-25 ml of whole seawater onto 0.2  $\mu$ M black polycarbonate filters, preserving with 0.5% glutaraldehyde and freezing at -20 °C. Cells were counted under epifluorescence miscroscopy, and cell size, shape and volume were used to calculate  $\mu$ g phytoplankton carbon per L. For heterotrophic bacteria, cyanobacteria, and picoeukaryote cell counts,

1.8 ml of whole seawater was preserved with a final concentration of 0.4% paraformaldehyde and flash frozen in liquid nitrogen. Samples were stored at -80 °C until analysis on a Beckman Coulter (Brea, CA) Altra flow cytometer, using 488 nm and UV nm laser excitation for simultaneous detection of DNA (Hoechst-stained cells, 1  $\mu$ g/ml final concentration), chlorophyll and phycoerythrin fluorescence, as well as forward and 90° light scatter (35).

#### Additional physical and biochemical data

Mooring data from Station M0 provided water column depth profiles of temperature and salinity and water velocities in the vertical and horizontal directions. Winds were measured from mooring Station M2 as described previously (50, 51).

#### Post-deployment extractions, qPCR and RT-qPCR

For the bench-top extractions and qPCR, DNA and RNA internal standards were added just prior to the initiation of cell lysis (see below) to control for differences in extraction efficiency between samples or extraction methods. The DNA internal standard consisted of *Thermus thermophilus* HB-8 genomic DNA (purchased from American Type Culture Collection), and the RNA internal standard consisted of reverse-transcribed mRNA from the T7 promoter to the NcoI restriction site of pFN18A HaloTag T7 Flexi Vector template (Promega, Madison, WI) (standard courtesy of Brandon Satinsky). Fluorogenic probe-based assays were designed for the L-Serine O-acetyltransferase (*sat1*) gene (Accession number AB159102; only ~15% overall sequence identity to other known

plant and bacterial *sat* genes, ref. 28) of the added *Thermus thermophilus* DNA standard and a 130 nt region of the HaloTag RNA standard (see below).

The DNA-only set of Niskin filters were extracted using the Qiagen DNAeasy modified protocol used in the ESP (42) except that DNA standard was added immediately prior to heat lysis. RNAlater-preserved ESP archive and Niskin-based filters (0.2  $\mu$ M and 5.0  $\mu$ M) were extracted for DNA and RNA using a modified protocol of the AllPrep DNA/RNA mini extraction kit (Qiagen) using bead-beating lysis. The AllPrep spin column (DNA) allows the flow-through to be applied to an RNeasy mini spin column (RNA) for simultaneous DNA and RNA extraction from the same filter. To counteract the low pH of the RNAlater in the lysate, the RLT Plus lysis buffer was amended with NaOH to obtain a final pH of  $\sim$ 7. A 250 ml 1:1 mix of 0.1 and 0.5 mm silica beads (BioSpec, Bartlesville, OK) and the DNA and RNA standards were added to the lysate immediately prior to bead beating. After bead-beating, the lysate was homogenized by passage through a 0.22 gauge needle, and the remainder of the extraction protocol followed the AllPrep manual descriptions for DNA and RNA. RNA extracts were DNase digested using the Turbo DNA-free kit (Ambion) with double the enzyme volume. RNA and DNA integrity was verified using an Agilent 2100 BioAnalyzer at the University of Georgia Genomics Facility.

Bench-top qPCR assays of DMSP genes for the ESP archive and Niskin filters used the same conditions and primer and probe concentrations as described for the ESP *in situ* qPCR assays (Table 4.1). The bench-top qPCR conditions for the Thermus DNA and HaloTag RNA standards are given in Table 4.1. All qPCR was carried out on an iCycler iQ or iCyler iQ5 (Bio-Rad, Hercules, CA). For reverse-transcription (RT)-qPCR, the

Invitrogen OneStep Express kit (SuperScript III) for specific priming of cDNA synthesis was used in 25.0 µl final volumes. qPCR analysis of a suite of DNA and RNA samples from Monterey Bay for which no internal standards were added confirmed that there was no amplification by the Thermus or HaloTag primer sets. There was also no crossreactivity of Roseo 2255 dmdA, Roseo 2255 dddP, and SAR11 D/1 dmdA primer sets with Thermus or HaloTag standards at  $10^7$ ,  $10^6$ , and  $10^5$  gene copies per reaction, or between Thermus and HaloTag primer sets and standards. Reactions without reverse transcriptase were run to confirm the absence of genomic DNA contamination. Triplicate no-template control reactions were included on every plate. DNA template was added in 1:10 or 1:100 dilutions, and RNA template was added in 1:10 or 1:20 dilutions. Ten-fold serially diluted standard curves representing 10<sup>1</sup> to 10<sup>7</sup> gene or mRNA copies per reaction were included on every plate. Bench-top standard curve efficiencies were 91%-105%, and linear regression  $r^2$  values were >0.98. Size and specificity of the qPCR and RT-qPCR products were verified by agarose gel electrophoresis. The limit of detection for quantification was 10 gene copies per reaction for both Roseo 2255 dmdA and dddP and 50-100 gene copies per reaction for SAR11 D/1 dmdA.

DNA extracted from archived ESP filters using the Allprep method (which yields both DNA and RNA) was highly variable, with copy numbers for DMSP genes differing by up to two orders of magnitude for closely spaced sample dates (Fig. 4.8A in Appendix C). qPCR quantification of internal standards in these samples was reproducible (Fig. 4.6), as was RNA recovery from them. A test with *T. thermophilus* DNA sheared to 100, 400, and 1000 bp fragments indicated that small DNA fragments were retained up to 8fold less efficiently than large fragments with the Allprep extraction method (Fig. 4.8B in Appendix C). Thus we hypothesize that DNA in ESP-stored filters was fragmented during storage or extraction, and DNA from these filters was subsequently used only to obtain relative community composition information from 16S rRNA gene amplifications. A second ESP archive filter collected on October 25 (Table 4.3 in Appendix C) was removed from analyses since both DNA and RNA yields and qPCR measurements were extremely low, resulting in 18 measurements of DMSP gene transcripts from ESP archived RNA samples.

Statistical analyses were carried out in R (45) using non-parametric Spearman's rank correlation coefficients and a significance level based on p-value cutoffs of <0.05. For correlation analyses, gene and cell count data were log transformed and all data were normalized to a mean of 0 and standard deviation of 1 (Z-score transformation). Environmental variables used in statistical analyses included measurements of CTD chlorophyll, salinity, temperature and % beam transmission, and discrete measurements of DMSPt, DMSPd, DMSPd consumption rate, DMSPd rate constant (k), DMSPd turnover as a fraction of DMSPt, Chl *a*, phaeopigments, DMSPp:Chl *a* ratio, heterotrophic bacteria, cyanobacteria, picoeukaryotes, dinoflagellates (*Prorocentrum micans*), diatoms (*Pseudonitzchia* sp.), and prymnesiophytes (coccolithophores).

For pairwise comparisons of gene abundance, a Wilcoxon rank sum test was used to test for significant differences in medians based on p-value cutoffs of <0.05.

#### 16S rRNA sequencing and analysis

ESP archive DNA samples from the 0.2-5.0  $\mu$ M and >5.0  $\mu$ M size fractions were used in triplicate PCR amplifications with the bacterial 16S rRNA primers (Bakt\_341F and

Bakt\_805R) from Herlemann et al. (17) over 25 cycles. The primers were modified with 454 Titanium adaptors and sample-specific 5-bp barcodes. PCR assays used the Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA) and 0.5 μM final concentrations of each primer. Following PCR, amplicons were purified with Agencourt Ampure XP (Beckman Coulter) using a 1:1 volume of PCR product to Ampure XP beads. Purified amplicons were quantified using Quant-iT PicoGreen (Invitrogen) and pooled in equal concentration and submitted to the Georgia Genomics Facility (University of Georgia) for Roche/454 Titanium sequencing.

16S rRNA sequences were analyzed using the QIIME (6) pipeline downloaded from http://www.qiime.org/. Sequences without perfect matches to primer and barcode sequences were removed, and remaining sequences were separated by barcode ID and denoised using AmpliconNoise (44). 16S rRNA sequences were clustered into operational taxonomic units (OTUs) based on a 97% sequence similarity and taxonomy was assigned using the Greengenes classifier using the latest build (gg\_otus\_4feb2011). Reference sequences from each OTU were also further compared to a marine 16S rRNA sequence custom database according to Biers et al. (3) using Smith-Waterman pairwise alignments (55) and requiring sequence overlaps of  $\geq$ 80%. If possible, sequences were assigned to species level taxa with  $\geq$ 97% identity across the overlap. Since the only available 16S rRNA sequence for HTCC2255 is a partial sequence [399 bp; (37)], a manual alignment was used to assess % identity over an ~80 bp overlap with amplicon reference sequences from OTUs of the order Rhodobacterales. *Thermus thermophilus* (DNA standard), Archaea, chloroplast, and unassigned (could not be classified to the

kingdom level; <2% of the total) sequences were removed prior to analyses of taxonomic structure.

## Results

#### Environmental dynamics

The autonomous remote sampling instrument, the Environmental Sample Processor (ESP), was moored in Monterey Bay within 0.2 mi of Station M0 (36.835 N, 121.901W) at a depth of ~8 m from September 28 - October 31, 2010. Winds were favorable for upwelling at the start of the sampling period (51) and coincided with a high Chl *a* concentration of ~15.3 ( $\pm$ 4.7) µg L<sup>-1</sup> and low water clarity from September 30 -October 6 (Fig. 4.1A and C; Fig. 4.2A), during which time salinity peaked at 33.77 and temperature steadily increased from 11.5 to 12.7 °C (Fig. 4.1B).

A reversal/relaxation of the upwelling-favorable winds occurred on October 5 and a change in hydrography was evident by October 7 as salinity abruptly decreased to ~33.6 and temperature increased to 13.5 °C (Fig. 4.1B and 4.2A). Depth profiles from Station M0 show this temporal shift throughout the water column (Fig. 4.2B and C). Strong vertical mixing ended by October 8 (Fig. 4.2D and E). A second shift from upwelling favorable-winds to relaxation/reversal occurred on October 12 - October 15, followed by a longer shift on October 19 - October 25, for which there was a subsequent two-day upwelling-favorable period, and then a final shift to relaxation/reversal on October 27 (Fig. 4.2A). These shifts typically coincided with increases in Chl *a* except for October 19 (Figs. 4.1 and 4.2).

Discrete measurements of DMSP concentration, consumption rates, and turnover rates were carried out on Niskin samples collected near the ESP on selected dates. DMSPp concentrations ranged from 5 nM to 378 nM while DMSPd concentrations were well below 5 nM except on October 11 (Fig. 4.3A). The highest co-occurring measurements of Chl a and DMSPp concentrations were observed on October 5, 12, and 29 (Fig. 4.3A), typically during shifts from upwelling favorable winds to relaxation/reversal of winds (Fig. 4.2A). The DMSP-producing dinoflagellate *Prorocentrum micans* (16) was one of the dominant phytoplankton groups at these time points (Fig. 4.9 in Appendix C). Throughout the deployment, DMSPp concentrations were significantly positively correlated with this DMSP-producer ( $\rho = 0.90, P < 0.05$ ) and not significantly correlated with any of the other phytoplankton groups measured (diatoms, cyanobacteria, or coccolithophores). In addition, the abundance of *P. micans*, DMSPp concentrations, DMSPt concentrations, Chl a concentrations, and heterotrophic bacteria were significantly correlated with one another (all Spearman's  $\rho = 0.73$ , P < 0.730.05).

Chl *a* measurements carried out on the discrete Niskin samples paralleled the trends observed for the continuous ESP CTD measurements (Figs. 4.1A and 4.3A). The DMSPp:Chl *a* ratio, a proxy for the relative contribution of DMSP to phytoplankton carbon content, was lowest at the end of September and increased continually through October (Fig. 4.3B). DMSP consumption rates varied from 2 nM day<sup>-1</sup> (September 29) to 56 nM day<sup>-1</sup> (October 29), and the rate constant for turnover of the DMSPd pool averaged 14.9 ( $\pm$  5.6) day<sup>-1</sup> (Fig. 4.3C). The fraction of the DMSPt pool turned over daily by the microbial community ranged from 0.06 (October 5) to 0.59 (September 30).

## Gene abundance

The abundance of the three bacterial DMSP degradation genes along with a SAR11 16S rRNA gene was measured by qPCR on the ESP from September 28 -October 27. Primers and probes used on the ESP were targeted to taxa that were abundant in a previous Monterey Bay metagenomic survey (46). For *dmdA*, the genes from Roseobacter clade strain HTCC2255-like cells (~24% of all metagenomic *dmdA* sequences), and SAR11 D/1-like cells (~19% of all metagenomic *dmdA* sequences) were examined, and for *dddP* the gene from Roseobacter clade strain HTCC2255-like cells (50% of all metagenomic *dddP* sequences) was examined. Because of configuration constraints on the ESP, in situ filters (those extracted and analyzed on the mooring) were used for DNA-based qPCR of whole water samples (>0.2 µm fraction). ESP archived filters (usually collected within 2 h of the *in situ* filters and stored on the ESP in RNAlater until analyzed in the laboratory) were used for RNA-based qPCR and were collected as two size fractions (free-living, 0.2 to 5.0 µm fraction; and particle associated,  $>5.0 \mu m$  fraction). Niskin samples were processed in the same way as the *in situ* ESP samples (n=11) or the ESP archived filters (n=6) (Table 4.3 in Appendix C). For any gene and transcript comparisons across sample types in the sections below, transcript measurements have been summed to represent whole water data (>0.2  $\mu$ m) unless specifically indicated.

For all genes, highest abundances occurred on October 5 (coinciding with peaks in Chl *a* and DMSPp concentration) and lowest abundance occurred in the time periods around September 29 for SAR11 D/1-like genes and October 20 for HTCC2255-like genes (Fig. 4.4A). The *in situ* abundance of the two HTCC2255-like genes was nearly

identical throughout the sampling period, averaging  $3.4 \times 10^7$  (*dmdA*) and  $3.1 \times 10^7$  (*dddP*) copies L<sup>-1</sup> (Fig. 4.4A), indicating that Monterey Bay populations of HTCC2255-like cells consistently harbor a copy of each gene. The *in situ* abundance of SAR11 D/1 *dmdA* was ~3-fold higher than the HTCC2255 genes, with an average of  $9 \times 10^7$  copies L<sup>-1</sup> (Fig. 4.4A), and was 9 - 22% of the SAR11 16S rRNA gene abundance (Fig. 4.4A) assuming one gene copy per cell. Gene abundance measures obtained from the discrete Niskin samples (processed as for the ESP *in situ* samples) were comparable to estimates from the ESP (Fig. 4.4A).

## Transcript abundance

DMSP transcript abundances in RNA extracted from ESP archived filters were positively correlated with the ESP *in situ* gene copies L<sup>-1</sup> (all Spearman's  $\rho \ge 0.63$ , P < 0.05), although transcript numbers were on average 1.5- to 2.5-fold more variable than gene copy numbers. HTCC2255 *dmdA* transcripts averaged 5.8x10<sup>5</sup> copies L<sup>-1</sup> (~60-fold lower than gene abundance) while HTCC2255 *dddP* transcripts averaged 5.1x10<sup>4</sup> copies L<sup>-1</sup> (~600-fold lower than gene abundance). SAR11 D/1 *dmdA* transcripts averaged 5.3x10<sup>5</sup> copies L<sup>-1</sup> (~170-fold lower than gene abundance) (Fig. 4.4B). All transcripts were in lowest abundance on September 28-29, and were highest on October 11-12 for HTCC2255 *dmdA* and on October 4-5 for both HTCC2255 *dddP* and SAR11 D/1 *dmdA*. Transcripts analyzed in each size fraction separately showed the free-living fraction contributed ~90% of the whole water signal for HTCC2255 transcripts and ~99.9% of the whole water signal for SAR11 transcripts (data not shown). Within the 5.0 µM fraction,

HTCC2255 *dmdA* transcripts were still >10-fold higher than HTCC2255 *dddP* transcripts.

On 6 dates, transcript abundance measures were also obtained from discrete, RNAlater-preserved Niskin samples, and these differed from the ESP archive data by less than 2-fold for all three genes (Fig. 4.4B).

#### *Expression ratios*

Gene expression ratios for the ESP samples (calculated as archived sample transcript copies: in situ sample gene copies) showed elevated expression levels for all genes during the week of October 11, and thereafter elevated expression ratios for both HTCC2255 *dmdA* and *dddP* on October 20 and for SAR11 D/1 *dmdA* on October 26 (Fig. 4.5A). Lowest expression ratios occurred on September 29 or 30 for all three genes (Fig. 4.5A). HTCC2255 *dmdA* gene expression ratios averaged 3.5-fold higher than SAR11 D/1 and 11.5-fold higher than HTCC2255 *dddP*.

Expression ratios calculated for the discrete Niskin samples were consistent with ESP data (Fig. 4.5B). We were also able to compare expression ratios in the two size fractions for the Niskin samples, since gene and transcript abundances were determined for both filter sizes (Fig. 4.5C). Expression ratios for HTCC2255 *dmdA* and *dddP* did not significantly vary by size fraction, but expression was significantly lower for SAR11 D/1 *dmdA* genes in the particle-associated fraction compared to the free-living fraction (Wilcoxon rank sum, P < 0.05).

#### Correlations with environmental data

DMSP gene data from the ESP *in situ* measurements, transcript abundance from ESP archive measurements, and ESP expression ratios were used in Spearman rank correlations with the environmental data measured *in situ* during the time-series or in discrete Niskin samples. *In situ* abundance of all DMSP genes measured on the ESP was positively correlated with numbers of heterotrophic bacteria ( $\rho \ge 0.78$ , P < 0.01 for all), Chl *a* ( $\rho \ge 0.69$ , P < 0.05), and salinity ( $\rho \ge 0.79$ , P < 0.01). The HTCC2255 *dmdA* was also positively correlated with DMSPp concentration ( $\rho = 0.58$ , P < 0.05), but while SAR11 D/1 *dmdA* and HTCC2255 *dddP* also appeared to be related to DMSPp concentration, these associations were not statistically significant ( $\rho \ge 0.55$ , P < 0.08 for both); this was due in part to different sets of dates with *in situ* qPCR data for these genes (Fig. 4.4A).

Transcript abundance of all DMSP genes was positively correlated with dinoflagellate abundance (*P. micans*) ( $\rho \ge 0.0.82$ , P < 0.05 for all) and negatively correlated with DMSPd turnover as a fraction of DMSPt ( $\rho \le -0.65$ , P < 0.05 for all), and these two variables were negatively correlated with one another ( $\rho$ =-0.89, P < 0.05). In addition, Roseobacter *dddP* transcripts were also significantly positively correlated with three other intercorrelated variables: DMSPt, DMSPp, and heterotroph cell number ( $\rho \ge 0.58$ , P < 0.05 for all); and SAR11 D/1 *dmdA* transcripts were also significantly positively correlated with DMSPt, DMSPp, Chl *a*, and DMSPd ( $\rho > 0.53$ , P < 0.05 for all) and negatively with % beam transmission ( $\rho = -0.5$ , P < 0.05). Gene expression ratios were positively correlated with DMSPp:Chl *a* ratios ( $\rho \ge 0.72$ , P < 0.05 for all) and temperature ( $\rho \ge 0.54$ , P < 0.05 for all).

#### Internal standard normalization

DNA and RNA internal standards were used in all laboratory-processed qPCRs; these allowed improved confidence in quantification by accounting for any between-sample differences in sample handling, extraction efficiencies, or thermal cycler performance. Recovery of the DNA standard, which consisted of genomic DNA from *Thermus thermophilus* HB-8 added to filters immediately prior to bench-top extraction, was determined by quantification of the *T. thermophilus sat1* gene. qPCR *sat1* recovery averaged ~18% of added copies, with no significant differences between extraction methods (Fig. 4.6). Recovery of the mRNA standard, which consisted of reverse-transcribed HaloTag vector added to filters immediately prior to bench-top extraction, was determined by quantification of an internal 130 nt region. RT-qPCR recoveries of the artificial mRNA averaged ~19% of added copies, with no significant differences between extractions except that the 5.0 µm RNA standard recovery from the Niskin samples was ~1.5 fold higher than the median (Fig. 4.6; Wilcoxon rank sum, *P*<0.05).

Transcript and gene copy counts were corrected based on internal standard losses for analyses that included an internal standard addition (DNA from Niskin filters, RNA from Niskin and ESP archived filters). For Niskin DNA samples, gene counts increased by ~4-fold following normalization, while for Niskin and ESP RNA samples, transcript counts increased by ~6 to 7-fold following normalization to the internal standard (Table 4.2). While normalization of gene and transcript counts increased estimates of absolute copy numbers in the seawater samples, abundance and expression patterns were not changed by this correction.

#### Bacterial community structure

The bacterial community composition was analyzed from 16S rRNA sequencing from the ESP archived filters (Fig. 4.10 in Appendix C). The 94,215 sequences obtained from both size fractions formed 4,712 OTUs at a >97% sequence identity. The community was dominated by Bacteroidetes (Flavobacteria) and Gammaproteobacteria at all sample dates and in both size fractions (Fig. 4.10 in Appendix C). The Rhodobacterales, which harbors the Roseobacter clade, made up ~17% (free-living) and ~8% (particle associated) of the 16S rRNA amplicons. The Rickettsiales, which harbors the SAR11, made up only ~2% and ~0.9% of the sequences in the size fractions (Fig. 4.10 in Appendix C). A single base mismatch in the reverse primer to >99% of available SAR11 16S rRNA sequences (including 7 cultured and 11,584 uncultured SAR11 representatives; Ribosomal Database Project, http://rdp.cme.msu.edu/ and the Gordon and Betty Moore Foundation Microbial Genome Sequencing Project,

http://camera.calit2.net/microgenome/) may have caused this unexpectedly low (by about 10-fold) representation.

A marine 16S rRNA gene database (3) was used to determine if any of the sequence clusters represented the bacterial taxa targeted with the qPCR probes. Four OTUs made up ~85% of the sequences in the Rhodobacterales, and one of these (accounting for half of all Roseobacter sequences) was 98.7% identical to the Roseobacter clade strain HTCC2255 16S rRNA gene. Three OTUs made up ~84% of the sequences in the SAR11 lineage, one of which was 100% identical to the *Candidatus* 'Pelagibacter ubique' HTCC7211 16S rRNA gene and two of which had insufficient similarity to classify them with any known SAR11 isolates. The HTCC7211-like cluster

is not likely to harbor a D/1 clade of SAR11 *dmdA* sequences, and the single SAR11 isolate that contains a clade D/1-like *dmdA* sequence (SAR11 alphaproteobacterium HIMB5) was not represented in the 16S rRNA amplicons. The inability to find the 16S rRNA cluster most likely responsible for the D/1 *dmdA* gene may reflect the primer mismatch to most SAR11 16S rRNA sequences.

## Discussion

In a coastal upwelling location such as Monterey Bay, bacterial turnover of DMSP likely occurs over short time scales, mediated by diverse DMSP degraders with varied degradation capabilities (29). We anticipated that a remote autonomous instrument capable of daily sampling and high resolution molecular measurements over a monthlong time frame would capture DMSP gene dynamics at unprecedented resolution with particular value for biogeochemical modeling.

Wind-driven relaxation shifts from upwelling periods occurred several times at the mooring Station M0 in Monterey Bay during the ESP deployment, and three of these coincided with measurements of high abundance of the DMSP-producing dinoflagellate *Prorocentrum micans* and elevated concentrations of DMSPp and Chl *a*. Heterotrophic bacteria (composed of Flavobacteria, Gammaproteobacteria, roseobacters, and SAR11, among others; Fig. 4.10 in Appendix C) were also elevated during these events, correlating with dinoflagellate cell numbers and DMSP and Chl *a* concentrations. To understand bacterially-mediated DMSP degradation within this temporally variable context, we measured gene abundance in real-time on the ESP (15 measurements within 29 days) and transcript abundance on samples collected by the ESP usually just before or

after the *in situ* sample and archived until retrieval (18 measurements within 30 days). In agreement with *in situ* gene monitoring and RNA archiving by the ESP in previous studies (38, 42, 48), qPCR measurements from Niskin bottle samples collected on 6 or 11 occasions (depending on the Niskin sample type) and processed immediately in the laboratory were similar to ESP measurements for both genes and transcripts. We took advantage of internal DNA and RNA standards, becoming more widely used in normalizing environmental qPCR and sequencing studies (10, 12, 65), to account for any differences in processing. Thus the data acquired remotely and obtained from autonomous sample archiving robustly characterized gene abundance and expression in high resolution during the course of several DMSP-producing events.

While some studies have speculated that specialized DMSP degraders become dominant when DMSP concentrations increase (69), the abundance of DMSP demethylase and cleavage genes in the Monterey Bay bacterial community tracked closely with total numbers of heterotrophic bacteria. Indeed, the overall consistency of DMSP gene abundance throughout much of the ocean surface waters (19, 66) suggests that DMSP cycling is mediated by widely distributed, non-specialist bacterial taxa. Under this scenario, transcript abundance rather than gene abundance would be a better indicator of conditions favorable for DMSP processing by bacterial taxa. Consistent with this idea, transcript numbers were several-fold more variable over time than gene copy numbers and transcripts of all three DMSP genes correlated with cell counts of the DMSP-producing dinoflagellate *P. micans*. Transcript numbers were between 60- to 600-fold lower than gene numbers, but the relatively low expression levels for DMSP degradation genes is consistent with what has been observed *in vitro* (S. Gifford,

*personal comm.*). The best correlate with gene expression ratios (transcripts:genes) for all genes was the DMSPp:Chl *a* ratio (Fig. 4.7), a parameter considered an indicator of the fractional importance of DMSP in the labile organic carbon pool (27, 45) and expected to track with the abundance of high DMSP-producing phytoplankton species. What noticeably was not correlated with expression ratio was the concentration or consumption rate of DMSPd. These findings are in line with previous studies of bacterial DMSP turnover using qPCR and environmental microarrays (30, 47) and support a central hypothesis regarding the factors that regulate bacterial DMSP degradation in ocean surface waters (41): that the proportional contribution of DMSP to the labile organic matter pool is an important regulatory signal for DMSP-degrading bacterioplankton.

In a comparison of the two bacterial taxa, Roseobacter clade strain HTCC2255 *dmdA* gene expression ratios were typically~3.5-fold higher than SAR11 D/1 *dmdA* ratios (Fig. 4.5). Malmstrom et al. (33) also reported high rates of Roseobacter DMSP assimilation on a per cell basis in comparison to other marine bacterial groups. On the other hand, SAR11 D/1 *dmdA* genes were 3-fold more abundant than Roseobacter genes, consistent with measurements in other marine surface waters (19, 30, 66, 68). Both groups therefore appear to be significant DMSP degraders in this system, with the HTCC2255-like populations more transcriptionally active on a per gene basis and the SAR11 D/1-like populations more abundant.

We asked whether there were differences between the two size fractions in expression of bacterial DMSP genes, particularly since roseobacters are often found in association with phytoplankton cells and other particulate material (4, 11, 20, 72) while

SAR11 cells are small and free-living (13, 36). For the HTCC2255-like genes, transcript abundance in the particle-associated fraction was ~10% of the whole water signal, while for SAR11 D/1 the particle associated fraction was <0.1%. Similarly, gene expression ratios in the particle-associated fraction were no different than for free-living in the case of the two HTCC2255-like genes, but for SAR11 D/1 the particle-associated gene expression ratio was significantly lower than for free-living. The SAR11-like cells harboring D/1 *dmdA* genes in Monterey Bay appear to play a less important role in DMSP degradation when associated with particulate material.

Transcript abundance cannot directly predict catalytic activity since posttranscriptional and post-translational regulation, enzyme efficiency and half-life, and substrate availability all affect final metabolic rates (1, 31, 34). Thus the consistently lower abundance of transcripts for *dddP* compared to *dmdA* in HTCC2255-like cells (averaging ~12-fold less) does not necessarily indicate that DMSP cleavage was occurring at a lower rate than demethylation. However, changes in the ratio of transcripts representing the two pathways (*dmdA:dddP*) may be informative regarding shifts in regulation within this Monterey Bay Roseobacter population. This ratio ranged from 5:1 on October 4 to 17:1 by October 11 (Fig. 4.11 in Appendix C), suggesting that DMSP cleavage was more important during the first DMSP-producing bloom (beginning on October 5) and that demethylation became more important after the bloom. Kiene et al. (25) proposed that DMSP degradation pathways are regulated by the C and S demand of the bacterial community, with bacteria switching to demethylation when more organic S is required by growing cells. Therefore, increases in *dmdA:dddP* expression ratios in bacteria such as Roseobacter clade strain HTCC2255 that possess both pathways could be a bioindicator of conditions unfavorable to DMS formation in ocean surface waters. Among the environmental parameters measured in this study, there were no significant correlates with the *dmdA:dddP* expression ratio; future high-resolution studies with an increased parameter set (e.g., bacterial production, bacterial C and S demand, solar radiation levels, etc.) may uncover candidate factors for the regulation of DMSP degradation pathways.

The high resolution of the ESP offered an unprecedented ability for remote sampling of DMSP-degrading marine bacteria, providing near-daily gene abundance and expression data over a month-long time period. In Monterey Bay, two abundant DMSPdegrading bacterial clades, one SAR11 and one Roseobacter, contributed hundreds of millions of DMSP-related genes per liter. The SAR11 D/1-like genes were more abundant but had low expression, particularly when present in the particle-attached fraction of the community. HTCC2255-like *dmdA* genes were less abundant but had higher per gene expression and both *dmdA* and *dddP* genes were expressed equally in particle-attached and free-living fractions. The proportional importance of DMSP in the labile organic matter pool, based on the DMSPp:Chl *a* ratio as a proxy for high DMSPproducing phytoplankton, was the best correlate with DMSP gene expression.

#### Acknowledgements

We thank J. Birch, B. Roman, and S. Jenson for technical expertise; E. Demir, K. Buck, F. Chavez, T. Pennington, M. Blum, J. Harvey, and C. Wahl for help with protocols and sampling; the Captain and crew of the R/V *Zephyr* and the R/V *John Martin*; B. Carter for assistance with MEGAMER facilities; K. Turk, M. Hogan, and B. Higgins for advice

on qPCR; L. Oswald for DMSP measurements; S. Tanner and R. Michisaki for phytoplankton enumeration; K. Selph for flow cytometry measurements; S. Bertilsson, D. Herlemann, and P. Yager for advice on 16S rRNA primers; S. Sharma, A. Rivers, and W. Sheldon for bioinformatics assistance; C. English for graphics help; and N. Levine for advice on manuscript preparation. This work was supported by grants from NSF (EnGen award OCE0724017) and the Gordon and Betty Moore Foundation, including the MEGAMER facility grant.

## References

- Alonso, C., and J. Pernthaler. 2006. Concentration-dependent patterns of leucine incorporation by coastal picoplankton. Appl. Environ. Microbiol. 72:2141-2147.
- Andreae, M. O. 1990. Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. Mar. Chem. 30:1-29.
- Biers, E. J., S. Sun, and E. C. Howard. 2009. Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. Appl. Environ. Microbiol. 75:2221-2229.
- Buchan, A., J. M. Gonzalez, and M. A. Moran. 2005. Overview of the marine Roseobacter lineage. Appl. Environ. Microbiol. 71:5665-5677.
- Burkill, P. H., S. D. Archer, C. Robinson, P. D. Nightingale, S. B. Groom, G.
  A. Tarran, and M. V. Zubkov. 2002. Dimethyl sulphide biogeochemistry within a coccolithophore bloom (DISCO): an overview. Deep-Sea Res. Pt II 49:2863-2885.
- Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7:335-336.
- Charlson, R. J., J. E. Lovelock, M. O. Andreae, and S. G. Warren. 1987.
  Oceanic phytoplankton, atmospheric sulfur, cloud albedo, and climate. Nature 326:655-661.
- Curson, A. R. J., R. Rogers, J. D. Todd, C. A. Brearley, and A. W. B.
  Johnston. 2008. Molecular genetic analysis of a dimethylsulfoniopropionate

lyase that liberates the climate-changing gas dimethylsulfide in several marine  $\alpha$ -proteobacteria and *Rhodobacter sphaeroides*. Environ. Microbiol. **10**:757-767.

- Del Valle, D. A., R. P. Kiene, and D. M. Karl. 2012. Effect of visible light on dimethylsulfoniopropionate assimilation and conversion to dimethylsulfide in the North Pacific Subtropical Gyre. Aquat. Microb. Ecol. 66:47-62.
- Devonshire, A., R. Elaswarapu, and C. Foy. 2011. Applicability of RNA standards for evaluating RT-qPCR assays and platforms. BMC Genomics 12:118.
- Geng, H., and R. Belas. 2010. Molecular mechanisms underlying roseobacterphytoplankton symbioses. Curr. Opin. Biotech. 21:332-338.
- Gifford, S. M., S. Sharma, J. M. Rinta-Kanto, and M. A. Moran. 2011.
  Quantitative analysis of a deeply sequenced marine microbial metatranscriptome.
  ISME J. 5:461-472.
- Giovannoni, S. J., H. J. Tripp, S. Givan, M. Podar, K. L. Vergin, D. Baptista,
  L. Bibbs, J. Eads, T. H. Richardson, M. Noordewier, M. S. Rappe, J. M.
  Short, J. C. Carrington, and E. J. Mathur. 2005. Genome streamlining in a
  cosmopolitan oceanic bacterium. Science 309:1242-1245.
- 14. Gonzalez, J. M., R. P. Kiene, and M. A. Moran. 1999. Transformation of sulfur compounds by an abundant lineage of marine bacteria in the α-subclass of the class *Proteobacteria*. Appl. Environ. Microbiol. 65:3810-3819.
- 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. Appl. Environ. Microbiol. 66:4237-4246.

- Hatton, A., and S. Wilson. 2007. Particulate dimethylsulphoxide and dimethylsulphoniopropionate in phytoplankton cultures and Scottish coastal waters. Aquat. Sci. 69:330-340.
- Herlemann, D. P. R., M. Labrenz, K. Jurgens, S. Bertilsson, J. J. Waniek, and A. F. Andersson. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 5:1571-1579.
- Howard, E. C., J. R. Henriksen, A. Buchan, C. R. Reisch, H. Bürgmann, R. Welsh, W. Y. Ye, J. M. Gonzalez, K. Mace, S. B. Joye, R. P. Kiene, W. B. Whitman, and M. A. Moran. 2006. Bacterial taxa that limit sulfur flux from the ocean. Science 314:649-652.
- Howard, E. C., S. Sun, E. J. Biers, and M. A. Moran. 2008. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. Environ. Microbiol. 10:2397-2410.
- Jasti, S., M. E. Sieracki, N. J. Poulton, M. W. Giewat, and J. N. Rooney-Varga. 2005. Phylogenetic diversity and specificity of bacteria closely associated with *Alexandrium* spp. and other phytoplankton. Appl. Environ. Microbiol. 71:3483-3494.
- Kiene, R. P., and G. Gerard. 1995. Evaluation of glycine betaine as an inhibitor of dissolved dimethylsulfoniopropionate degradation in coastal waters. Mar. Ecol-Prog. Ser. 128:121-131.
- 22. Kiene, R. P., and L. J. Linn. 2000. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico. Limnol. Oceanogr. 45:849-861.

- Xiene, R. P., and L. J. Linn. 2000. The fate of dissolved dimethylsulfoniopropionate (DMSP) in seawater: tracer studies using <sup>35</sup>S-DMSP. Geochim. Cosmochim. Ac. 64:2797-2810.
- Kiene, R. P., L. J. Linn, and J. A. Bruton. 2000. New and important roles for DMSP in marine microbial communities. J. Sea Res. 43:209-224.
- Kiene, R. P., L. J. Linn, J. Gonzalez, M. A. Moran, and J. A. Bruton. 1999. Dimethylsulfoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. Appl. Environ. Microbiol. 65:4549-4558.
- Kiene, R. P., and D. Slezak. 2006. Low dissolved DMSP concentrations in seawater revealed by small-volume gravity filtration and dialysis sampling. Limnol. Oceanogr. Methods 4:80-95.
- Kirst, G. O. 1990. Salinity tolerance of eukaryotic marine algae. Annu. Rev. of Plant Phys. 41:21-53.
- Kobayashi, S.-i., R. Masui, S. Yokoyama, S. Kuramitsu, and H. Takagi. 2004.
  A novel metal-activated L-serine O-acetyltransferase from *Thermus thermophilus* HB8. J. Biochem. 136:629-634.
- 29. Ledyard, K. M., and J. W. H. Dacey. 1996. Kinetics of DMSP lyase activity in coastal seawater, p 325-335, In R.P. Kiene, P.T. Visscher, M.D. Keller, and G.O. Kirst (eds) Biological and environmental chemistry of DMSP and related sulfonium compounds, Plenum Press, New York.
- Levine, N. M., V. A. Varaljay, D. Toole, A., J. W. H. Dacey, S. C. Doney, and
  M. A. Moran. 2012. Environmental, biochemical, and genetic drivers of DMSP
degradation and DMS production in the Sargasso Sea. Environ. Microbiol. doi: 10.1111/j.1462-2920.2012.02700.x.

- 31. Maier, T., A. Schmidt, M. Guell, S. Kuhner, A.-C. Gavin, R. Aebersold, and L. Serrano. 2011. Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol. Syst. Biol. 7:511.
- Malmstrom, R. R., R. P. Kiene, M. T. Cottrell, and D. L. Kirchman. 2004. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic Ocean. Appl. Environ. Microbiol. 70:4129-4135.
- 33. Malmstrom, R. R., R. P. Kiene, and D. L. Kirchman. 2004. Identification and enumeration of bacteria assimilating dimethylsulfoniopropionate (DMSP) in the North Atlantic and Gulf of Mexico. Limnol. Oceanogr. 49:597-606.
- Mata, J., S. Marguerat, and J. Bähler. 2005. Post-transcriptional control of gene expression: a genome-wide perspective. Trends Biochem. Sci. 30:506-514.
- Monger, B. C., and M. R. Landry. 1993. Flow cytometric analysis of marine bacteria with hoechst 33342. Appl. Environ. Microbiol. 59:905-911.
- 36. 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.
- 37. Newton, R. J., L. E. Griffin, K. M. Bowles, C. Meile, S. Gifford, et al. 2010.
   Genome characteristics of a generalist marine bacterial lineage. ISME J. 4:784-798.

- 38. Ottesen, E. A., R. Marin, III, C. M. Preston, C. R. Young, J. P. Ryan, C. A. Scholin, and E. F. DeLong. 2011. Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. ISME J. 5:1881-1895.
- Pennington, J. T., and F. P. Chavez. 2000. Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989-1996 in Monterey Bay, California. Deep-Sea Res. Pt. II 47:947-973.
- 40. Pichereau, V., J.-A. Pocard, J. Hamelin, C. Blanco, and T. Bernard. 1998. Differential effects of dimethylsulfoniopropionate, dimethylsulfonioacetate, and other s-methylated compounds on the growth of *Sinorhizobium meliloti* at low and high osmolarities. Appl. Environ. Microbiol. 64:1420-1429.
- Pinhassi, J., R. Simó, J. M. González, M. Vila, L. Alonso-Sáez, R. P. Kiene, M. A. Moran, and C. Pedrós-Alió. 2005. Dimethylsulfoniopropionate turnover is linked to the composition and dynamics of the bacterioplankton assemblage during a microcosm phytoplankton bloom. Appl. Environ. Microbiol. 71:7650-7660.
- 42. Preston, C. M., A. Harris, J. P. Ryan, B. Roman, R. Marin, III, S. Jensen, C. Everlove, J. Birch, J. M. Dzenitis, D. Pargett, M. Adachi, K. Turk, J. P. Zehr, and C. A. Scholin. 2011. Underwater application of quantitative PCR on an ocean mooring. PLoS ONE 6:e22522.
- Preston, C. M., R. Marin III, S. D. Jensen, J. Feldman, J. M. Birch, E. I.
   Massion, E. F. DeLong, M. Suzuki, and K. S. Wheeler, C.A. Scholin. 2009.
   Near real-time, autonomous detection of marine bacterioplankton on a coastal

mooring in Monterey Bay, California, using rRNA-targeted DNA probes. Environ. Microbiol. **11**:1168-1180.

- 44. Quince, C., A. Lanzen, R. Davenport, and P. Turnbaugh. 2011. Removing noise from pyrosequenced amplicons. BMC Bioinformatics **12**:38.
- R Development Core Team. 2011. R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna Austria. ISBN 3-900051-07-0. http://www.R-project.org.
- 46. Rich, V. I., V. D. Pham, J. Eppley, Y. Shi, and E. F. DeLong. 2011. Timeseries analyses of Monterey Bay coastal microbial picoplankton using a 'genome proxy' microarray. Environ. Microbiol. 13:116-134.
- 47. Rinta-Kanto, J. M., H. Bürgmann, S. M. Gifford, S. Sun, S. Sharma, D. A. del Valle, R. P. Kiene, and M. A. Moran. 2011. Analysis of sulfur-related transcription by Roseobacter communities using a taxon-specific functional gene microarray. Environ. Microbiol. 13:453-467.
- 48. Robidart, J. C., C. M. Preston, R. W. Paerl, K. A. Turk, A. C. Mosier, C. A. Francis, C. A. Scholin, and J. P. Zehr. 2011. Seasonal *Synechococcus* and *Thaumarchaeal* population dynamics examined with high resolution with remote in situ instrumentation. ISME J. 6:513-523.
- Ryan, J. P., J. F. R. Gower, S. A. King, W. P. Bissett, A. M. Fischer, R. M. Kudela, Z. Kolber, F. Mazzillo, E. V. Rienecker, and F. P. Chavez. 2008. A coastal ocean extreme bloom incubator. Geophys. Res. Lett. 35:L12602.
- So. Ryan, J. P., D. I. Greenfield, R. Marin III, C. Preston, B. Roman, S. Jensen,
  D. Pargett, J. Birch, C. Mikulski, G. J. Doucette, and C. A. Scholin. 2011.

Harmful phytoplankton ecology studies using an autonomous molecular analytical and ocean observing network. Limnol. Oceanogr. **56**:1255-1272.

- St. Ryan, J. P., M. A. McManus, R. M. Kudela, M. L. Artigas, J. G. Bellingham,
  F. P. Chavez, G. J. Doucette, D. Foley, M. Godin, J. B. J. Harvey, R. Marin
  III, M. Messie, C. Mikulski, J. T. Pennington, F. Py, K. Rajan, I. Shulman, Z.
  Wang, and Y. Zhang. Submitted. Boundary influences on phytoplankton
  ecology in a stratification-enhanced upwelling shadow. Deep-Sea Res. Pt. II.
- 52. Scholin, C. A., G. J. Doucette, S. Jensen, B. Roman, D. Pargett, R. Marin III, et al. 2009. Remote detection of marine microbes, small invertebrates, harmful algae, and biotoxins using the Environmental Sample Processor (ESP). Oceanography 22:158-167.
- Shenoy, D. M., and J. S. Patil. 2003. Temporal variations in dimethylsulphoniopropionate and dimethyl sulphide in the Zuari estuary, Goa (India). Mar. Environ. Res. 56:387-402.
- 54. Slezak, D., R. Kiene, D. Toole, R. Simó, and D. Kieber. 2007. Effects of solar radiation on the fate of dissolved DMSP and conversion to DMS in seawater. Aquat. Sci. 69:377-393.
- Smith, T. F., and M. S. Waterman. 1981. Identification of common molecular subsequences. J. Mol. Biol. 147:195-197.
- 56. Sunda, W., D. J. Kieber, R. P. Kiene, and S. Huntsman. 2002. An antioxidant function for DMSP and DMS in marine algae. Nature 418:317-320.
- 57. Suzuki, M. T., C. M. Preston, O. Béjà, J. R. de la Torre, G. F. Steward, and
   E. F. DeLong. 2004. Phylogenetic screening of ribosomal RNA gene-containing

clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. Microb. Ecol. **48**:473-488.

- 58. Suzuki, M. T., C. M. Preston, F. P. Chavez, and E. F. DeLong. 2001. Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay. Aquat. Microb. Ecol. 24:117-127.
- 59. Todd, J. D., A. J. Curson, C. L. Dupont, P. Nicholson, and A. W. B. Johnston. 2009. The *dddP* gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. Environ. Microbiol. **11**:1624-1625.
- Todd, J. D., A. R. J. Curson, M. Kirkwood, M. J. Sullivan, R. T. Green, and
   A. W. B. Johnston. 2011. DddQ, a novel, cupin-containing,
   dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine
   bacteria. Environ. Microbiol. 13:427-438.
- Todd, J. D., M. Kirkwood, S. Newton-Payne, and A. W. B. Johnston. 2012.
   DddW, a third DMSP lyase in a model Roseobacter marine bacterium, *Ruegeria* pomeroyi DSS-3. ISME J. 6:223-226.
- Todd, J. D., R. Rogers, Y. G. Li, M. Wexler, P. L. Bond, L. Sun, A. R. J. Curson, G. Malin, M. Steinke, and A. W. B. Johnston. 2007. Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. Science 315:666-669.

- 63. Turner, S. M., G. Malin, P. S. Liss, D. S. Harbour, and P. M. Holligan. 1988. The seasonal variation of dimethyl sulfide and dimethylsulfoniopropionate concentrations in nearshore waters. Limnol. Oceanogr. 33:364-375.
- 64. Vairavamurthy, A., M. O. Andreae, and R. L. Iverson. 1985. Biosynthesis of dimethylsulfide and dimethylpropiothetin by *Hymenomonas carterae* in relation to sulfur source and salinity variations. Limnol. Oceanogr. **30**:59-70.
- 65. van Doorn, R., M. M. Klerks, M. P. E. van Gent-Pelzer, A. G. C. L. Speksnijder, G. A. Kowalchuk, and C. D. Schoen. 2009. Accurate quantification of microorganisms in PCR-inhibiting environmental DNA extracts by a novel internal amplification control approach using Biotrove OpenArrays. Appl. Environ. Microbiol. **75**:7253-7260.
- Varaljay, V. A., S. M. Gifford, S. T. Wilson, S. Sharma, D. M. Karl, and M.
   A. Moran. 2012. Bacterial dimethylsulfoniopropionate degradation genes in the oligotrophic North Pacific Subtropical Gyre. Appl. Environ. Microbiol. 78:2775-2782.
- 67. Varaljay, V. A., E. C. Howard, S. Sun, and M. A. Moran. 2010. Deep sequencing of a dimethylsulfoniopropionate-degrading gene (*dmdA*) by using PCR primer pairs designed on the basis of marine metagenomic data. Appl. Environ. Microbiol. **76**:609-617.
- Vila-Costa, M., J. M. Rinta-Kanto, S. Sun, S. Sharma, R. Poretsky, and M.
   A. Moran. 2010. Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. ISME J. 4:1410-1420.

- 69. Vila, M., R. Simo, R. P. Kiene, J. Pinhassi, J. M. Gonzalez, M. A. Moran, and C. Pedros-Alia. 2004. Use of microautoradiography combined with fluorescence in situ hybridization to determine dimethylsulfoniopropionate incorporation by marine bacterioplankton taxa. Appl. Environ. Microbiol. 70:4648-4657.
- Wolfe, G. V., M. Steinke, and G. O. Kirst. 1997. Grazing-activated chemical defence in a unicellular marine alga. Nature 387:894-897.
- Yoch, D. C. 2002. Dimethylsulfoniopropionate: its sources, role in the marine food web, and biological degradation to dimethylsulfide. Appl. Environ.
   Microbiol. 68:5804-5815.
- 72. Zubkov, M. V., B. M. Fuchs, S. D. Archer, R. P. Kiene, R. Amann, and P. H. Burkill. 2001. Linking the composition of bacterioplankton to rapid turnover of dissolved dimethylsulphoniopropionate in an algal bloom in the North Sea. Environ. Microbiol. 3:304-311.
- 73. Zubkov, M. V., B. M. Fuchs, S. D. Archer, R. P. Kiene, R. Amann, and P. H. Burkill. 2002. Rapid turnover of dissolved DMS and DMSP by defined bacterioplankton communities in the stratified euphotic zone of the North Sea. Deep-Sea Res. Pt II 49:3017-3038.

Table 4.1. Primer and	probe sequences,	concentrations, and	l annealing tem	peratures used in c	PCR and RT-q	PCR
		,	<u> </u>			

Primer name	Forward	Conc (nM)	Reverse	Conc (nM)	Probe	Anneal temp °C	Amplicon size (bp)
HTCC2255 dddP	TGGTGGCGAATGGATAGAAAC	1,000	GGGCCACATTCTTGAAACCAG	1,500	TGTCCTTGGTCCGGTTGTTAATAA GCG	53	74
HTCC2255 dmdA	GGGCGAATGTTGAATATAGCAAGA	1,000	CCGGACCAAGCTGAGAGT	1,500	TTGGTCGAAACAGGGTGGATTTGA AA	53	83
SAR11 D/1 dmdA	AGATGTTATTATTGTCCAATAATTG ATG	1,500	ATCCACCATCTATCTTCAGCTA	500	ACAGGATCATTAACTAAATTTCC	49	89
SAR11 16S*	CTCTTTCGTCGGGGAAGAAA	500	CCACCTACGWGCTCTTTAAGC	1,500	TTACCGCGGCTGCTGGCAC	59	155
Thermus	CTCCGACTACGTCAACGTCTAC	500	GCTTCCGATGATGGTCTCCTTC	1,500	AGCCACACCCACCACGTCCTGG	62	87
HaloTag	GCGCTGGTCGAAGAATACAT	1,000	TTGCAGTTAGGCAGGCTTTT	300	ACTGGCTGCACCAGTCCCCT	61	130

\*from Suzuki et al.(58)

Gene	Niskin DN	leasy DNA	Niskin All	lprep RNA	ESP Allprep RNA		
	Measured Normalized		Measured	Normalized	Measured	Normalized	
HTCC2255 dmdA	3.5x10 <sup>7</sup>	1.4x10 <sup>8</sup>	1.4x10 <sup>6</sup>	8.1x10 <sup>6</sup>	5.8x10 <sup>5</sup>	$4 x 10^{6}$	
HTCC2255 dddP	$3.1 \times 10^7$	1.3x10 <sup>8</sup>	$4.3 \text{x} 10^4$	2.6x10 <sup>5</sup>	5.1x10 <sup>4</sup>	3.5x10 <sup>5</sup>	
SAR11 D/1 dmdA	$1.6 \times 10^7$	6.2x10 <sup>8</sup>	6.1x10 <sup>5</sup>	$3.7 \times 10^{6}$	5.3x10 <sup>5</sup>	3.6x10 <sup>6</sup>	

Table 4.2. Measured versus normalized (corrected for internal standard recovery) gene or transcript counts per liter. Data are averages for all dates and each extraction method (n=11 for Niskin DNeasy DNA, n=6 for Niskin Allprep RNA, n=18 for ESP Allprep RNA)

Figure 4.1. ESP CTD time-series data collected at 8.1 m ( $\pm$  0.7) near Station M0 in Monterey Bay from September 28 - October 31, 2010. A) Chl *a* (black line) with discrete Niskin-based Chl *a* measurements also shown (gray diamonds). B) Temperature (gray line); Salinity (black line); C) Percent beam transmission.





Figure 4.2. Regional wind forcing from Station M2 (modified from ref. 51) and depth profiles of environmental data from Station M0 in Monterey Bay from September 28 - October 31, 2010. In panel A, shaded periods indicate relaxation/reversal of winds and unshaded periods indicate when winds were upwelling favorable. For panels B-E, the dotted lines at 8.0 m indicate the average depth of the ESP. The solid circles along the dotted line in panel B indicate dates when DMSP measurements were made. A) Temperature; B) Salinity; C) and D) Water velocity in the x and y directions, respectively.



Station M2 regional wind forcing (September - October 2010)

Figure 4.3. DMSP measurements collected at 9.2 m ( $\pm$  0.7) at the deployed ESP (near Station M0) from September 28 - October 29, 2010. Error bars are standard deviations of biological replicates.



Date (September - October 2010)

Figure 4.4. Gene and transcript copies L<sup>-1</sup> collected near Station M0 from September 28 -October 29, 2010 either by ESP or Niskin collection methods. At the top is a diagram of known DMSP degradation pathways based on HTCC2255 and SAR11 genome sequences. A) ESP *in situ* (circles) and Niskin (triangles) gene abundance and B) ESP archived samples (circles) and Niskin (squares) transcript abundance. The color of the symbol corresponds to the gene group: SAR11 16S rRNA (orange), SAR11 D/1 *dmdA* (green), HTCC2255 *dmdA* (blue) and HTCC2255 *dddP* (red). Gray error bars are standard deviations for Niskin replicate samples.



Date (Sept-Oct 2010)

Figure 4.5. Gene expression ratios for ESP and Niskin samples collected near Station M0 from September 28 - October 29, 2010. Expression ratios are shown for ESP samples (A), Niskin samples for both size fractions combined (B) and Niskin samples for each size fraction separately (C). The color of the bar corresponds to the gene group: HTCC2255 *dmdA* (blue), HTCC2255 *dddP* (red), and SAR11 D/1 *dmdA* (green). Expression was below detection limit for quantification in the >5.0  $\mu$ m fraction on October 27 for HTCC2255 *dddP* and on October 13 and 27 for SAR11 D/1 *dmdA*. Propagated error bars are standard deviations for Niskin replicate RNA and DNA samples.



Figure 4.6. DNA and RNA internal standard recoveries using qPCR.  $1.2 \times 10^7$  or  $9.8 \times 10^6$  DNA internal standard copies were added to DNA filters and  $4.2 \times 10^6$  RNA internal standard copies were added to RNA filters. The box is the interquartile range (IQR) of the data, the black bar indicates the median, and the whiskers are the maximum and minimum data values within 1.5 times the IQR. Outliers (> 1.5 times the IQR) are shown as open circles.



Figure 4.7. Scatter plots of DMSPp:Chl *a* ratios versus gene expression ratios. Spearman rank correlation coefficients are shown in the top left of each panel. Note the break in the X-axis for October 20 and October 26 samples; these samples were included in calculations of correlation coefficients.



SAR11 Clade D/1 dmdA



HTCC2255 dddP



HTCC2255 dmdA

## CHAPTER 5

## CONCLUSIONS

Environmental sequence data and PCR-based measurements of DMSP demethylation and cleavage genes were used here to better understand the bacterial role in sulfur cycling in natural marine systems. Marine metagenomic data served as the basis for the design of *dmdA* and *dddP* primer sets for three marine environments studied. The first study focused on the abundant and diverse DMSP demethylase gene (*dmdA*) following a comprehensive environmental primer design procedure, high-throughput sequencing of PCR amplicons, and cluster-based diversity assessment of sequence variation. The next two studies provided the first looks into spatial and temporal patterns in gene abundance and expression of diverse *dmdA* and *dddP* groups in both open-ocean and coastal ocean environments. The goals of these studies were to determine the composition of bacterial DMSP degraders in each system and relate gene-based patterns of abundance and expression with environmental variables to identify potential drivers of DMSP fate.

In the *in silico* bioinformatic approach described in Chapter 2, *dmdA* primers were designed from marine metagenomic data from the 2007 Global Ocean Sampling (GOS) metagenomic surface water survey, and resulting amplicon pools were deeply sequenced. Environmental sequence data from the GOS were mined to optimize primer design for *dmdA* genes in an iterative bioinformatic pipeline, and when applied to a southeastern U.S. coastal bacterioplankton community, the primers revealed more than 700 clusters (at

90% sequence similarity) from taxonomic groups affiliated with SAR11, Roseobacter, and Gammaproteobacteria lineages, along with other as-yet unidentified groups. The primers captured >17,000 unique amplicon sequences while retaining specificity for the target *dmdA* group.

The optimized *dmdA* primer sets were also used to assess differences between particle-associated and free-living size-fractions of these coastal bacterial communities, as these represent microscale niches that differ in concentration and supply of DMSP for bacterial taxa. Differences in cluster richness were observed for some subclades between size fractions, although most of the *dmdA* clusters were shared between the two. This suite of metagenomically-optimized and environmentally-relevant primer sets were also appropriate for quantitative PCR (qPCR) approaches, and in subsequent studies were used for tracking temporal and spatial dynamics of gene abundance and expression.

In Chapter 3, primers for dmdA and dddP genes were quantified over a 10 month time period in the North Pacific Subtropical Gyre at Station ALOHA, as part of the Hawaii Ocean Time-series (HOT) long-term oceanographic observatory. The abundance of eight dmdA subclades using primers designed in Chapter 2 and a dddP subclade (designed to target Roseobacter-like genes) were temporally stable but spatially distinct between the two discrete depths measured, the nutrient-deplete upper euphotic zone at 25 m and the nutrient-replete lower euphotic zone at ~100 m (the Deep Chlorophyll Maximum or DCM). Gene abundances were higher at 25 m (~13% of cells harbored a dmdA gene) than at the DCM (~6.5%), with some subclades primarily confined to one or the other depth, and correlated significantly with environmental parameters such as DMSP and DMS concentrations, temperature, and photosynthetically active radiation.

Intra-depth analyses provided insight into relationships of some *dmdA* and *dddP* gene groups with phytoplankton pigments, such as those diagnostic of diatoms, which signified a possible interaction between the bacteria harboring DMSP degradation genes and specific phytoplankton taxa.

SAR11 *dmdA* genes dominated the gene pool at Station ALOHA, as *dmdA:dddP* gene ratios were ~9:1 and 80% of the *dmdA* sequences were SAR11-like genes. However, SAR11 *dmdA* gene expression from two subclades was low, on the order of 1:350 to 1:1,400 transcripts per gene copy, which could be due to an artifact of the low nucleic acid yields for RNA analysis or could indicate an uncoupling of transcription and biogeochemical rates of demethylation. A survey of metagenomic and metatranscriptomic data from this same site supported our PCR-based estimates of the gene composition and abundance, and also confirmed that expression levels of DMSP genes were relatively low. The stable and abundant gene frequencies suggested that DMSP degradation is not carried out by specialized DMSP-degraders, but instead is mediated by widespread bacterial taxa that have maintained the genetic capability to take advantage of this important molecule. Based on the composition of the gene pool, we hypothesized that DMSP degradation is dominated year-round by SAR11 bacteria in the upper ocean, with lesser but consistent involvement of other bacterial taxa such as the Roseobacter and Gammaproteobacteria groups.

In Chapter 4, bacterial degradation dynamics were measured on a more highlyresolved time scale in Monterey Bay, a coastal upwelling system in which temporal phytoplankton abundance dynamics, and therefore DMSP concentrations, are highly variable. In order to capture bacterial DMSP-degrading genes in the context of these

dynamics, we used primers designed from Monterey Bay metagenomic data on a moored autonomous remote instrument ('ESP') capable of near-daily qPCR analysis. The measurements of genes designed based on Roseobacter clade strain HTCC2255 *dmdA* and *dddP* sequences and SAR11 Clade D/1 *dmdA* sequences were made *in situ* on the remote instrument, and transcripts from these genes were quantified on samples collected concurrently but archived on the instrument for the duration of the month-long deployment. The high-resolution method proved two insights: a) changes in wind shifts coincided with at least 3 major shifts in phytoplankton abundance (dinoflagellates), DMSP concentrations, with changes in DMSP gene abundance and expression and b) gene abundance and expression measurements from the remotely collected samples were validated by comparison with traditional Niskin bottle samples and recoveries of internal gene and transcript standards.

Transcript abundances for all genes were correlated with DMSP-producing dinoflagellate cell counts and gene expression ratios were correlated with particulate (within phytoplankton) DMSP per Chl *a*, which suggests that expression is linked to the abundance of high-DMSP-producing phytoplankton taxa. HTCC2255 *dmdA* genes were less abundant, but had higher expression ratios than SAR11 D/1. SAR11 *dmdA* gene expression ratios were also significantly different between the free-living and particle-associated size fractions, in contrast to Roseobacter genes for which there was no apparent difference. These results may relate to the different niches occupied by SAR11 (free-living) and Roseobacter (often phytoplankton-associated) cells. Roseobacter *dmdA* and *dddP* genes showed differences in peak expression dynamics, suggesting that expression ratios of the genes from relatives of this species could be used as an *in situ* 

bioassay for the importance of demethylation vs. cleavage during DMSP degradation. On the whole, both SAR11 and Roseobacter members are significant contributors to DMSP transformations in Monterey Bay.

Molecular tools for tracking diverse and widespread genes mediating bacterial DMSP degradation were developed and tested, and the resultant data are helping to build an improved foundation for exploring interactions between bacteria, phytoplankton taxa, and DMSP pools in regulating cycling of DMSP-derived carbon and sulfur in marine environments.

## APPENDIX A

## CHAPTER 2 SUPPLEMENTARY MATERIAL

<sup>&</sup>lt;sup>1</sup> Supporting Online Material for: Varaljay, V.A., E.C. Howard, S. Sun, and M.A. Moran. 2010. Applied and Environmental Microbiology. 76: 609-617. Reprinted here with permission of publisher.

Table 2.4. DNA samples from Sapelo Island, GA used in *dmdA* analysis. All samples were collected from surface water (0.5 m depth).

Site	Size fraction (µm)	Collection date	# samples		
	0.2 - 1.0	Fall 2000	2		
Dean Creek	8.0 - 1.0	Fall 2000	2		
Doboy Sound	0.2 - 1.0	Fall 2000	2		
Dobby Sound	8.0 - 1.0	Fall 2000	2		
Dean Creek	0.2 - 1.0	Winter 2000	2		
Dean Creek	8.0 - 1.0	Winter 2000	2		
Doboy Sound	0.2 - 1.0	Winter 2000	2		
Dobby Sound	8.0 - 1.0	Winter 2000	2		
Dean Creek	0.2 - 1.0	Spring 2001	2		
Dean Creek	8.0 - 1.0	Spring 2001	2		
Doboy Sound	0.2 - 1.0	Spring 2001	2		
Doboy Sound Dean Creek Doboy Sound Dean Creek	8.0 - 1.0	Spring 2001	2		
Deen Creek	0.2 - 1.0	Summer 2001	2		
Deall Creek	8.0 - 1.0	Summer 2001	2		
	0.2 - 1.0	Summer 2001	2		
Doboy Sound	8.0 - 1.0	Summer 2001	2		
	0.2 - 1.0	Fall 2003	2		
Dean Creek	8.0 - 1.0	Fall 2003	2		
	0.2 - 1.0	Fall 2003	2		
Doboy Sound	8.0 - 1.0	Fall 2003	2		
	0.2 - 1.0	Spring 2004	2		
Dean Creek	8.0 - 1.0	Spring 2004	2		
D -1 0 1	0.2 - 1.0	Spring 2004	2		
Doboy Sound	8.0 - 1.0	Spring 2004	2		
	0.2 - 1.0	Summer 2004	2		
Dean Creek	8.0 - 1.0	Summer 2004	2		
	02 - 10	Summer 2004	2		
Doboy Sound	8.0 - 1.0	Summer 2004	2		
	0.2 - 1.0	Fall 2004	2		
Dean Creek	8.0 - 1.0	Fall 2004	2		
	02 - 10	Fall 2004	2		
Doboy Sound	8.0 - 1.0	Fall 2004	2		
	0.2 1.0	Winter 2005	2		
Dean Creek	0.2 - 1.0 8 0 - 1 0	Winter 2005	$\frac{1}{2}$		
	0.0 - 1.0	Winter 2005	2		
Doboy Sound	0.2 - 1.0 8.0 - 1.0	Winter 2005	2		
	0.0 - 1.0	Spring 2005	$\frac{2}{2}$		
Dean Creek	0.2 - 1.0 8 0 - 1 0	Spring 2005	2		
	0.0 - 1.0	Spring 2005	4		

Table 2.5. Additional GOS-based *dmdA* primer pairs tested *in silico*. These primer pairs either targeted very few GOS *dmdA* sequences or did not amplify *dmdA* from SIMO environmental DNA.

				No.	No.	No. (%) target GOS reads binding primers			
Primer name	Primer version	<i>dmdA</i> position	Amplicon length (bp)	Primer sequence	target GOS reads	target GOS reads in range	$\leq 4 \leq 6$ mismatches mismatches		Notes
A/3	specific	572-757	185	A/3 sp FP - CTCGCTCAGGTTTCTCAAAA A/3 sp RP – TACGGCTGGTCATATCATTG	15	10	4 (40%)	4 (40%)	No <i>dmdA</i> amplification
B/1	specific	338-452	114	B/I sp FP - GAGATGATAAATTCTGGCTTTCTGTGGC B/I sp RP - AAAGGTGAAACATCTGGCTCAGACA	13	13	12 (92%)	12 (92%)	No <i>dmdA</i> amplification
B/2	specific	208-378	170	B/2 sp FP - GATGCCCAGAGATTGATACAATTGA B/2 sp RP - AACCCAGCATTTTCATCTATTAATGGA	6	5	5 (100%)	5 (100%)	No <i>dmdA</i> amplification
C/1	degenera te	125-320	195	C/I dg FP – TTGAAGAWGATTAYTGGCA C/I dg RP – CAGAACAGGATCATTDATXAT	97	44	6 (14%)	18 (41%)	Targeted too few GOS reads <sup>a</sup>
D/2	specific	269-377	108	D/2 sp FP - GGTGTTATTATTGCCCACTCATTG D/2 sp RP -	N/A	N/A	N/A	N/A	Targeted too few GOS reads
E/2	specific	80-154	74	E/1 sp FP - GCATCGGTTTACAATCATCTAGTC E/1 sp RP – GCTCACGCAAATGCCAATAG	3	3	2 (67%)	2 (67%)	No <i>dmdA</i> amplification

<sup>a</sup>  $\overline{C/1}$  also targeted a significant percentage of non-target GOS *dmdA* sequences

Table 2.6. BLASTx and clustering results for *dmdA* amplicons of the particle-associated size fraction from southeastern U.S. coastal seawater.

Primer name	Clade	Subclade	% correct clade(s)	% correct subclade (of correct clade)	% incorrect clade	% not dmdA <sup>a</sup>	No. sequences resampled	Normalized no. <i>dmdA</i> clusters <sup>b</sup>
dmdAU	All	All	81 <sup>c</sup>	n/a	n/a	19	400	68
A/1-sp	Clade A	Subclade 1	98.9	99.8	0.5	0.6	2500	33
A/2-sp	Clade A	Subclade 2	99.2	98.8	0.2	0.6	3500	18
A/2-dg	Clade A	Subclade 2	96.5	98.8	0.9	2.6	3500	28*
A/2-ino	Clade A	Subclade 2	99.6	99.4	< 0.1	0.4	3500	19
B/3-sp	Clade B	Subclade 3	98.6	98.0	1.0	0.4	5500	34
B/4-sp	Clade B	Subclade 4	33.3	99.8	66.0	0.7	1500	19
C/2-sp	Clade C	Subclade 2	75.0	77.5	22.0	3.0	1200	33
C/2-dg	Clade C	Subclade 2	37.0	83.7	60.5	2.5	1200	45*
C/2-ino	Clade C	Subclade 2	36.4	64.0	61	2.6	1200	25*
D/1-sp	Clade D	Subclade 1	90.3	98.1	0.4	9.3	6000	173
D/3-sp	Clade D	Subclade 3	99.5	90.6	0.1	0.4	4300	28
D/3-dg	Clade D	Subclade 3	98.0	98.0	2.0	< 0.1	4300	29
D/all-sp	Clade D	All	99.43	n/a	0.15	0.42	4500	83
D/all-dg	Clade D	All	99.9	n/a	< 0.01	0.1	4500	82
D/all-ino	Clade D	All	99.2	n/a	0	0.8	4500	101*
E/2-sp	Clade E	Subclade 2	95.1	95.1	1.7	3.2	3000	43
E/2-dg	Clade E	Subclade 2	98.5	98.5	0.6	0.9	3000	35*

<sup>a</sup>Includes sequences with hits to gcvT and those with no hits.<sup>b</sup>Average of 1000 resamplings (see Methods) using indicated population sizes. Cluster numbers marked with an asterisk (\*) were significantly different (p < 0.05) from that obtained by the specific version of that primer pair.

<sup>c</sup> Clade distributions for *dmdA* sequences captured by the universal primer pair in the particle-associated size fraction are as follows: Clade A, 36.6%; Clade B, 1.5%; Clade C, 7.0%; Clade D, 43.4%; Clade E, 5.5%; and unclassified, 6.0%.

Figure 2.4. Protocol for metagenomic primer design and bioinformatic pipeline.



Figure 2.5. Three-way Venn diagram of unique and shared *dmdA* clusters (A) and percent of sequences present in those unique and shared clusters (B) amplified from coastal bacterioplankton using specific, degenerate, and inosine versions of the Clade D primer pair (D/all-sp, D/all-dg, and D/all-ino).


# APPENDIX B

## CHAPTER 3 SUPPLEMENTARY MATERIAL

<sup>&</sup>lt;sup>1</sup> Supporting Online Material for:

Varaljay, V.A., S.M. Gifford, S.T. Wilson, S. Sharma, D.M. Karl, M.A. Moran. 2012. Applied and Environmental Microbiology. 78:2775-2782. Reprinted here with permission of publisher.

Cruise	Collection Date(s)	Depths sampled (m)
HOT201	05/27/08 / 05/28/08	25, 130 / 25, 140
HOT202	06/26/08	25, 125
HOT203	07/26/08	25, 125
HOT204	08/17/08	25, 125
HOT205	10/10/08	25
HOT206	11/30/08	25, 110
HOT208	01/21/09	25, 125
HOT209	02/17/09	25, 125

Table 3.1. HOT DNA and RNA sample collection

Query sequence	Accession numbers and organisms
DmdA	AAV95190 (Ruegeria pomerovi DSS-3)
	EAQ43549 (Roseobacter sp. MED193)
	EAU51309 (Alpha proteobacterium HTCC2255)
	YP 265671 (Pelagibacter ubique HTCC1062)
	ZP_05069448 (Candidatus Pelagibacter sp. HTCC7211)
	ZP_05070099 (Candidatus Pelagibacter sp. HTCC7211)
	YP_003550401.1 (Candidatus Puniceispirillum marinum
	IMCC1322)
	ZP_01625100 (Marine gamma proteobacterium HTCC2080)
DddP	ZP_00959238 (Roseovarius nubinhibens ISM)
	EAU51270 (Alpha proteobacterium HTCC2255)
	YP_167522 (Ruegeria pomeroyi DSS-3)
DddQ	ZP_00960431 (Roseovarius nubinhibens ISM)
	ZP_00960430 (Roseovarius nubinhibens ISM)
	YP_166837.1 (Ruegeria pomeroyi DSS-3)
	ZP_01741685.1 ( <i>Rhodobacterales bacterium</i> HTCC2150)
	ZP_05079306 (Rhodobacterales bacterium Y41)
	ZP_01756428 ( <i>Roseobacter</i> sp. SK209-2-6)
	<u>ZP_05340956.1 (<i>Thalassiobium</i> sp. R2A62)</u>
DddD	YP_166942.1 ( <i>Ruegeria pomeroyi</i> DSS-3)
	<u>ZP_01598691 (Marinomonas MWYL1)</u>
DddL	ZP_00955343.1 (Sulfitobacter sp. EE-36)
	ZP_00998135 (Oceanicola batensis H1CC2597)
DddY	ADT64689 (Alcaligenes faecalis)
	YP_003654286 and YP_003656922.1 (Arcobacter nitrofigilis
	DSM 7299)
	YP_001048659.1 (Shewanella baltica OS155)
	ABI73720.1 (Shewanella frigidimarina NCIMB 400)
	YP_001672568.1 (Shewanella halifaxensis HAW-EB4)
	YP_001503782.1 (Shewanella pealeana ATCC 700345)
	YP_002309652.1 (Shewanella piezotolerans WP3)
	YP_001185042.1 (Shewanella putrefaciens CN-32)
	YP_001/586/3.1 (Shewanella woodyi A1CC 51908)
	<u>YP_/34281.1 and YP_/382/3.1 (Snewanella sp. MR-4)</u>
DddW	YP_165/16.1 ( <i>Ruegeria pomeroyi</i> DSS-3)
	ZP_0105/82/.1 ( <i>Roseobacter</i> sp. MED193)
Proteorhodopsin	ZP_01264205.1 ( <i>Candidatus</i> Pelagibacter ubique HTCC1002)
	YP_002591282.1 (Candidatus Pelagibacter sp. HTCC7211)
	ZP_01447408.1 ( <i>Rhodobacterales bacterium</i> HTCC2255)
	ZP_01236264.1 (Vibrio angustum S14)
	ZP_02194911.1 (Vibrio campbellii AND4)

Table 3.2. Accession numbers for query sequences used in all BLAST analyses

	YP_266049.1 (Candidatus Pelagibacter ubique HTCC1062)
	ZP_01223638.1 (Gamma proteobacterium HTCC2207)
	ZP_03701022.1 (Flavobacteria bacterium MS024-3C)
	ZP_03702117.1 (Flavobacteria bacterium MS024-2A)
	ZP_01616930.1 (Marine gamma proteobacterium HTCC2143)
	ZP_01161099.1 (Photobacterium sp. SKA34)
	ZP_01117885.1 (Polaribacter irgensii 23-P)
	YP_001277280.1 ( <i>Roseiflexus</i> sp. RS-1)
	ZP_01253360.1 (Psychroflexus torquis ATCC 700755)
	ZP_01734914.1 (Flavobacteria bacterium BAL38)
	YP_001813902.1 (Exiguobacterium sibiricum 255-15)
	ZP_01049273.1 (Dokdonia donghaensis MED134)
	YP_644794.1 (Rubrobacter xylanophilus DSM 9941)
	YP_002678278.1 (Octadecabacter antarcticus 307)
RecA	P0A7G6.2 (Escherichia coli K-12)
16S rRNA	16S rrsE ribosomal RNA locus b4007 (Escherichia coli str. K-
	12 substr. MG1655)

Figure 3.7. Biogeochemical data at Stn ALOHA for 25 m (open symbols) and the DCM (closed symbols). Data collected on HOT cruises 201-209.



Figure 3.8. Scatter plots of gene data and environmental parameters. (A) dddP gene abundance versus fucoxanthin (ng L<sup>-1</sup>), (B) A/2 (circles) and E/2 (triangles) versus fucoxanthin (ng L<sup>-1</sup>), (C) A/1 versus Chl *a* (µg L<sup>-1</sup>), and (D) dmdA versus temperature (°C). Spearman rank correlation coefficients are indicated in the bottom left hand corner in each plot.









Figure 3.9. Percent representation of *dmdA* (black bars) and proteorhodopsin (gray bars) genes in metagenomic and metatranscriptomic datasets (refs. 1, 2) collected at Stn ALOHA in March 2006 (HOT179). The percent representation was calculated as the number of hits retrieved from BLAST divided by the database size and normalized by gene length\*100.



## **Appendix B References**

- Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. DeLong. 2008. Microbial community gene expression in ocean surface waters. P. Natl. Acad. Sci. USA. 105:3805-3810.
- 2. Shi, Y., G. W. Tyson, and E. F. DeLong. 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. Nature **459**:266-269.

# APPENDIX C

## CHAPTER 4 SUPPLEMENTARY MATERIAL

<sup>&</sup>lt;sup>1</sup> Supporting Material for: Varaljay, V.A., J. Robidart, C.M. Preston, J.P. Ryan, R. Marin III, R.P. Kiene, J.P. Zehr. C.A. Scholin, and M.A. Moran. To be submitted to Environmental Microbiology.

Table 4.3.	ESP- a	and Niskin	-based co	ollection	dates a	ind times	for mole	ecular a	nd bioch	emical	samples of	during	Septemb	er and Oc	tober
2010.															

ESP in situ	ESP archives	Niskin (matches ESP <i>in situ</i>	Niskin (matches ESP archive	Niskin supplementary biochemical measurements <sup>¥</sup>
0/00/0010 1 51 DM	0/00/0010 10 10 DM	extraction method)	extraction method)	
9/28/2010 1:51 PM	9/28/2010 12:12 PM			9/28/2010 11:20 AM <sup>2,0</sup>
	9/28/2010 3:26 PM			9/29/2010 2:00 PM <sup>a,e</sup>
9/29/2010 12:18 PM	9/29/2010 1:44 PM			
9/30/2010 9:05 AM	9/30/2010 10:46 AM	9/30/2010 12:09 PM		9/30/2010 12:09 PM <sup>a</sup>
	10/4/2010 1:20 PM			
10/5/2010 8:36 AM	10/5/2010 9:51 AM		10/5/2010 9:50 AM	10/5/2010 9:50 AM <sup>a,b,c</sup>
10/6/2010 9:05 AM	10/6/2010 10:28 AM			10/6/2010 9:11:00 AM <sup>a,b</sup>
10/11/2010 10:16 AM	10/11/2010 12:33 PM	10/11/2010 11:37 AM	10/11/2010 11:37 AM	10/11/2010 11:37 AM <sup>b</sup>
10/12/2010 9:05 AM	10/12/2010 10:54 AM	10/12/2010 7:06 AM		10/12/2010 7:06 AM <sup>a,b,c</sup>
10/13/2010 9:05 AM	10/13/2010 10:50 AM	10/13/2010 8:43 AM	10/13/2010 8:43 AM	10/13/2010 8:43 AM <sup>b</sup>
10/14/2010 9:05 AM	10/14/2010 10:55 AM	10/14/2010 8:09 AM		10/14/2010 8:09 AM <sup>a,b</sup>
10/18/2010 9:05 AM	10/18/2010 10:50 AM	10/18/2010 2:10 PM		10/18/2010 2:10 PM <sup>b,c</sup>
		10/19/2010 9:27 AM		10/19/2010 9:27 AM
10/20/2010 9:56 AM	10/20/2010 11:26 AM	10/20/2010 8:44 AM		10/20/2010 8:44 AM <sup>a,b,c</sup>
10/21/2010 9:05 AM	10/21/2010 10:46 AM		10/21/10 9:17 AM	10/21/10 9:17 AM <sup>a,b</sup>
10/25/2010 3:06 PM	10/25/2010 9:06 AM	10/25/2010 1:43 PM		10/25/2010 1:43 PM <sup>b</sup>
	10/25/2010 7:28 PM*			
10/26/2010 9:06 AM	10/26/2010 10:53 AM	10/26/2010 9:05 AM		10/26/2010 9:05 AM <sup>a,b,c</sup>
10/27/2010 9:06 AM	10/27/2010 10:48 AM		10/27/2010 7:51 AM	10/27/2010 7:51 AM <sup>b</sup>
	10/28/2010 1:05 PM	10/28/2010 7:55 AM		10/28/2010 7:55 AM <sup>a,b</sup>
			10/29/2010 10:05 AM	10/29/2010 10:05 AM <sup>b,c</sup>

<sup>\*</sup>Superscripts are shown for dates when <sup>a</sup>DMSP consumption, <sup>b</sup>bacteria and picoeukarote cell counts, or <sup>c</sup> phytoplankton measurements were conducted. \*Not included in analyses

Figure 4.8. Gene copies L<sup>-1</sup> for ESP and Niskin samples collected near Station M0 from September 28 - October 29, 2010 and DNA recovery as a function of molecule size. A) ESP archive (circles) and Niskin (squares) gene abundance. The color of the symbol corresponds to the gene group: SAR11 D/1 dmdA (green), HTCC2255 dmdA (blue) and HTCC2255 dddP (red). Gray error bars are standard deviations for Niskin biological replicate samples. B) Box and whisker plots of recovery of DNA standards sheared to different sizes and extracted using the AllPrep method. The box is the interquartile range (IQR) of the data, the black bar indicates the median, and the whiskers are the maximum and minimum data values within 1.5 times the IQR.





Figure 4.9. Phytoplankton taxa in samples collected on 8 days near Station M0 between September 28 and October 29, 2010.



Figure 4.10. Bacterial composition based on 16S rRNA sequences amplified from DNA obtained from the ESP archived samples. A) 0.2-5.0  $\mu$ m and B) >5.0  $\mu$ m fractions. Asterisks indicate archive samples collected by the ESP on the same day but separated by 3 hours.



Date (September - October 2010)

Figure 4.11. HTCC2255 *dmdA:dddP* transcript ratio. Arrows indicate the lowest (5:1) and highest (17:1) ratios.

