The ISME Journal (2009) 3, 364–373

www.nature.com/ismej

ORIGINAL ARTICLE

Gene transfer agent (GTA) genes reveal diverse and dynamic *Roseobacter* and *Rhodobacter* populations in the Chesapeake Bay

Yanlin Zhao[1,2], Kui Wang[1], Charles Budinoff[3], Alison Buchan[3], Andrew Lang[4], Nianzhi Jiao[2] and Feng Chen[1]

[1]*Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, MD, USA;* [2]*State Key Laboratory of Marine Environmental Sciences, Xiamen University, Xiamen, China;* [3]*Department of Microbiology, University of Tennessee, Knoxville, TN, USA and* [4]*Department of Biology, Memorial University of Newfoundland, St John's, Newfoundland, Canada*

Within the bacterial class Alphaproteobacteria, the order Rhodobacterales contains the *Roseobacter* and *Rhodobacter* clades. Roseobacters are abundant and play important biogeochemical roles in marine environments. *Roseobacter* and *Rhodobacter* genomes contain a conserved gene transfer agent (GTA) gene cluster, and GTA-mediated gene transfer has been observed in these groups of bacteria. In this study, we investigated the genetic diversity of these two groups in Chesapeake Bay surface waters using a specific PCR primer set targeting the conserved Rhodobacterales GTA major capsid protein gene (*g5*). The *g5* gene was successfully amplified from 26 Rhodobacterales isolates and the bay microbial communities using this primer set. Four *g5* clone libraries were constructed from microbial assemblages representing different regions and seasons of the bay and yielded diverse sequences. In total, 12 distinct *g5* clusters could be identified among 158 Chesapeake Bay clones, 11 fall within the *Roseobacter* clade, and one falls in the *Rhodobacter* clade. The vast majority of the clusters (10 out of 12) lack cultivated representatives. The composition of *g5* sequences varied dramatically along the bay during the wintertime, and a distinct *Roseobacter* population composition between winter and summer was observed. The congruence between *g5* and 16S rRNA gene phylogenies indicates that *g5* may serve as a useful genetic marker to investigate diversity and abundance of *Roseobacter* and *Rhodobacter* in natural environments. The presence of the *g5* gene in the natural populations of *Roseobacter* and *Rhodobacter* implies that genetic exchange through GTA transduction could be an important mechanism for maintaining the metabolic flexibility of these groups of bacteria.

The ISME Journal (2009) **3,** 364–373; doi:10.1038/ismej.2008.115; published online 20 November 2008
Subject Category: microbial ecology and functional diversity of natural habitats
Keywords: *Roseobacter*; *Rhodobacter*; GTA; diversity; Chesapeake Bay

Introduction

The *Roseobacter* clade can comprise upwards of 25% of the marine bacterioplankton and consists of at least 41 major phylogenetic clusters (in total more than 141 phylogenetic clusters) based on ⩾99% 16S rRNA gene sequence similarity (Buchan *et al.*, 2005; Wagner-Döbler and Biebl, 2006). Unlike most other numerically abundant marine bacterial lineages, strains of *Roseobacter* are readily cultivated and this clade, as a whole, is well represented by cultured isolates. Among the 141 recognized *Roseobacter* clusters, 120 clusters (85%) contain cultivable representatives (Buchan *et al.*, 2005). The studied members of the *Roseobacter* group contain diverse specific biological and ecological features and can consume various carbon and sulfur compounds such as dimethyl sulfoniopropionate (DMSP) (González *et al.*, 1999; Miller and Belas, 2004) and carbon monoxide (Moran *et al.*, 2004; Tolli *et al.*, 2006). Thus, they are important to global

Correspondence: F Chen, Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 E Pratt Street, Baltimore, MD 21202, USA.
E-mail: chenf@umbi.umd.edu or NZ Jiao, State Key Laboratory of Marine Environmental Sciences, Xiamen University, Xiamen, China.
E-mail: jiao@xmu.edu.cn
Received 7 August 2008; revised 11 October 2008; accepted 20 October 2008; published online 20 November 2008

biogeochemical carbon and sulfur cycles. Isolation and molecular ecology studies have revealed that the *Roseobacter* clade occupies diverse marine environments but is most predominant in coastal waters (Buchan *et al.*, 2005). Furthermore, many roseobacters have been found in association with algal blooms (González *et al.*, 2000; Alavi *et al.*, 2001; West *et al.*, 2008), and some roseobacters have been isolated from, and found to be dominant in polar environments (Brown and Bowman, 2001; Brinkmeyer *et al.*, 2003; Selje *et al.*, 2004; Prabagaran *et al.*, 2007). The first complete genome sequence of a marine roseobacter, *Silicibacter pomeroyi*, was reported by Moran *et al.* (2004), and there are now complete or draft genome sequences for ca. 40 marine roseobacters, representing diverse clusters (Brinkhoff *et al.*, 2008). This information will facilitate studies of these important organisms at the genomic level.

Bacteria in the *Rhodobacter* clade occur more frequently in freshwater or estuarine environments (Crump *et al.*, 1999, 2004; Kan *et al.*, 2007), and can also occur in high abundance in some marine environments (Hiraishi and Ueda, 1995). Similar to *Roseobacter*, members of *Rhodobacter* are also metabolically versatile and exhibit genomic complexity (Nereng and Kaplan, 1999; Choudhary *et al.*, 2004, 2007).

An unusual bacteriophage-like vehicle of genetic exchange known as a gene transfer agent (GTA) was first reported in the purple non-sulfur bacterium *Rhodobacter capsulatus* (Marrs, 1974). GTA is a small phage-like particle released by the bacteria, and each particle contains a random ca. 4.5-kb fragment of bacterial genomic DNA (Solioz and Marrs, 1977) that can be transferred between cells (Solioz *et al.*, 1975; Biers *et al.*, 2008). This has made GTA a very valuable tool for the study of *R. capsulatus* for more than 30 years, aiding with the construction of new strains (for example, Scolnik and Haselkorn, 1984; Lilburn *et al.*, 1992; Lang and Beatty, 2001, 2002) and mapping of genetic loci (for example, Wall *et al.*, 1975; Yen and Marrs, 1976; Wall and Braddock, 1984).

The *R. capsulatus* GTA (RcGTA) has been genetically characterized by Lang and Beatty (2000). The RcGTA gene cluster comprises 15 genes, and recently, conserved GTA gene clusters were found in other species and appear to be limited to the Rhodobacterales and some other Alphaproteobacteria (Lang and Beatty, 2007; Biers *et al.*, 2008). Currently, all but one of the available Rhodobacterales genomes contain a conserved set of genes that constitute the RcGTA gene cluster (Lang and Beatty, 2007; Biers *et al.*, 2008; Paul, 2008), but these genes are not found in other major bacterioplankton lineages, such as *Prochlorococcus*, *Synechococcus* and *Pelagibacter* spp. GTA-related gene transfer was suggested as a potential adaptation mechanism for these bacteria to maintain the metabolic flexibility to the changing marine environment (Biers *et al.*, 2008). Among the GTA genes, the major capsid gene (*g5*) is highly conserved among all of these bacteria and the phylogeny based on *g*5 is consistent with that based on 16S rRNA genes (Lang and Beatty, 2007).

Chesapeake Bay is the largest estuary in North America. It receives a great amount of freshwater and nutrient inputs from tributary rivers. Salinity, nutrient and other environmental factors display significant spatial and seasonal variation in the bay (Kan *et al.*, 2006). Similarly, the bacterioplankton in the bay exhibit strong distribution patterns (Kan *et al.*, 2006). Members of the class Alphaproteobacteria, particularly *Roseobacter*, are one of the dominant bacterial populations in Chesapeake Bay and their sequences can constitute more than a third of 16S rRNA gene clone libraries (Kan *et al.*, 2007). In a recent study of the bay, 11 *Roseobacter* clusters were identified based on bacterial rRNA operon clone library analysis, and greater genetic diversity of the *Roseobacter* clade was found in winter than in summer (Kan *et al.*, 2008). Notably, several novel clusters of *Roseobacter* appeared to be unique to Chesapeake Bay (Kan *et al.*, 2008). Members of the *Rhodobacter* (related to *Pseudorhodobacter* spp.) were detected in the upper bay during the wintertime, but were not detected in the summer (Kan *et al.*, 2007).

In this study, we developed a primer set to target the *Roseobacter* and *Rhodobacter* GTA *g5* gene to explore (1) whether we can detect *Roseobacter* and *Rhodobacter* GTA major capsid genes in the natural aquatic environment; (2) the diversity of *g5* gene sequences in *Roseobacter* and *Rhodobacter* and (3) the spatial and temporal genetic variations of g5 gene diversity in Chesapeake Bay.

Materials and methods

Sample collection and microbial DNA preparation
The four Chesapeake Bay samples used in this study included three samples collected from stations 908, 818 and 707 (Supplementary Figure S1) in March 2003, and one sample collected from station 818 in July 2003. Stations 908, 818 and 707 in March 2003 were chosen to represent the upper, middle and lower bay in winter, whereas station 818 in July 2003 represents a middle bay sample in summer (Table 1). The winter DNA samples used in this study were the same as those used for earlier 16S rRNA gene clone library analyses (Kan *et al.*, 2007, 2008). The sample collection and DNA preparation methods were described in detail by Kan *et al.* (2006). Briefly, 500 ml sea water was filtered onto 0.2-μm-pore-size polycarbonate filters (47 mm diameter; Millipore, Billerica, MA, USA) and microbial community DNA was extracted from the filters using the phenol–chloroform protocol as described earlier (Kan *et al.*, 2006).

Bacterial counts, chlorophyll a *and nutrient data*
The biological and hydrological data in Table 1 were acquired from the existing database of the

Table 1 Characteristics of environmental parameters and clone information for each sampling station

Parameters	*Upper bay (station 908) March 2003*	*Middle bay (station 818) March 2003*	*Lower bay (station 707) March 2003*	*Middle bay (station 818) July 2003*
Location	39°08′N, 76°20′W	38°18′N, 76°17′W	37°07′N, 76°07′W	38°18′N, 76°17′W
Sampling date	March 2003	March 2003	March 2003	July 2003
Water temperature (°C)	1.2	1.8	4.4	23.8
Salinity (p.p.t.)	10	15.8	23	8.9
Chlorophyll *a* ($\mu g\,l^{-1}$)	41.6	22.5	14.9	19
NH_4 (μM)	1.15	0.59		10
NO_3+NO_2 (μM)	42.0	17.6	2.83	13.8
PO_4 (μM)	0.58	0.48	0.33	0.65
Bacterial abundance (10^6 cells ml^{-1})[a]	1.24 ± 0.51	0.57 ± 0.2	0.45 ± 0.17	4.35 ± 1.6
Sequenced clone number	44	42	42	30
OTU number	15	11	9	8
Clusters	8	6	8	5
Coverage (%)	84.1	88.0	92.9	83.3
Shannon–Weaver index (*H′*)	2.385	2.066	1.619	1.482
Simpson (*D*)	0.884	0.850	0.727	0.680
Evenness	0.724	0.715	0.561	0.550

Abbreviation: OTU, operational taxonomic unit.
[a]Mean ± standard deviation.

Microbial Observatories of Virioplankton Ecology (MOVE) project (http://www.virusecology.org/MOVE/Home.html), and the methods for measuring these parameters have been described elsewhere (Kan *et al.*, 2006, 2007).

Isolation of Roseobacter *and* Rhodobacter

Roseobacter strains isolated from Chesapeake Bay in July 2007 were obtained by direct plating of sea water onto agar plates containing 500 μM DMSP as the sole carbon source. The growth medium contained 200 mM NaCl, 50 mM $MgSO_4$, 5 mM $CaCl_2$, 150 μM K_2HPO_4, 8 mM NH_4Cl, 5 mM Tris-HCl (pH 7.6). Filter-sterilized vitamin, trace metal and iron stock solutions were added to this basal media as described earlier (Budinoff and Hollibaugh, 2007) and these along with DMSP were added to autoclaved salt solutions. Agar plates were prepared with purified agar at 0.8% w/v and kept aerobically in a temperature-controlled incubator (20 °C). Alternatively, several *Roseobacter* and *Rhodobacter* strains were obtained from Chesapeake Bay in April 2008 and Virginia Beach in March 2007 by using the previously described DMSP enrichment method (González *et al.*, 2003) with some modification. Sea water samples were diluted with sterilized artificial sea water, then amended with 100 μM DMSP. After 1 week of incubation (20 °C, in the dark), 10 μl of the enrichment culture was plated on low-nutrient sea water medium plates (González *et al.*, 1999) containing 100 μM DMSP and kept at 20 °C temperature in the dark. Colonies were randomly selected, purified and identified by 16S rRNA gene analysis as described earlier (Giovannoni, 1991).

Primer design and PCR amplification

Degenerate PCR primers specific for the *Roseobacter* and *Rhodobacter* clades of the order Rhodobacterales were designed based on an alignment of the GTA capsid protein gene (*g5*) sequences from 13 members of the Rhodobacterales, the genome sequences of which were known (Supplementary Figure S2). The conserved amino-acid domains GY/FLVDPQT and AKPHVLF (corresponding to amino-acid residues 109–116 and 362–368 for *S. pomeroyi* DSS-3, respectively) were selected for design of the forward primer MCP-109F, 5′-GGCT A(T/C)CTGGT(G/C)GATCC(G/C)CA(G/A)AC-3′ and reverse primer MCP-368R, 5′-TAGAACAG(G/C) AC(G/A)TG(G/C)GG(T/C)TT(G/T)GC-3′, respectively. All the PCR reactions were performed in 50 μl volume containing 1 × reaction buffer (Genescript, Scotch Plains, NJ, USA) with 1.5 mM $MgCl_2$, 100 μM of each deoxynucleoside triphosphate, 10 pmol of each primer and 1 U *Taq* DNA polymerase (Genescript). For *Roseobacter* isolates, DNA released from boiled exponentially growing cultures or extracted using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, USA) were used as templates. To ensure the quality and quantity of DNA inputs for PCR amplification of environmental samples, extracted bacterial community DNA samples were first amplified with GenomiPhi V2 (GE Healthcare, Piscataway, NJ, USA) according to the manufacturer's protocol and ca. 100 ng DNA was subsequently used as a template for each PCR reaction. The PCR program included an initial denaturing step at 94 °C for 3 min, followed by 35 cycles of 94 °C for 1 min, annealing at 58 °C for 30 s, and 72 °C for 1 min, and followed by a final extension step at 72 °C for 10 min.

Cloning, sequencing and phylogenetic analysis
The purified *g5* amplicons were cloned with the TOPO-TA cloning kit and ligated plasmids were transformed into TOP10 competent *Escherichia coli* cells according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). Transformants were selected on Luria–Bertani agar plates containing ampicillin (50 μg ml^{-1}) and X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) (20 μg ml^{-1}). White colonies were screened by PCR using the vector primers M13F (5′-GTAAAACGACGGCCAG-3′) and T7R (5′-TAATACGACTCACTATAGGG-3′). Purified PCR products were then sequenced using an automated sequencer ABI310 (Applied Biosystems, Foster City, CA, USA) in the Biological and Analytical Laboratory at the Center of Marine Biotechnology, University of Maryland Biotechnology Institute. For each library, 30–44 clones with appropriately sized inserts were sequenced.

All of the *g5* sequences were edited using Mac Vector 7.1 program (GCG) to remove plasmid vector and primer sequences, and the DNA sequences were subsequently translated into amino-acid sequence. The resulting capsid protein sequences obtained in this study were aligned and compared with reference sequences in the database. Neighbor-joining phylogenetic trees were constructed using MEGA 4.0 software (Tamura *et al.*, 2007). For translated amino-acid sequence of g5 gene, the evolution distances were calculated under the Jones–Taylor–Thornton model with rate variation among sites and complete deletion of gaps. For DNA sequences of the *g5* and 16S rRNA genes, the evolutionary distances were calculated using maximum composite likelihood with complete deletion of gaps.

Statistical analyses
Rarefaction analysis was performed using Analytic Rarefaction version 1.3 (http://www.uga.edu/~strata/software). The coverage for each clone library (C, the fraction of the population represented by the phylotypes that have been discovered in each clone library) was calculated by the equation $C=1-(n/N)\times 100$, where n is the number of unique clones and N is the total number of clones examined (Ravenschlag *et al.*, 1999). Operational taxonomic unit was defined as greater than or equal to 98% identical at the DNA sequence level. The statistical methods used for the estimation of species richness and diversity indices were based on coverage. Coverage-based estimations of species richness and evenness as well as the Shannon–Weaver index (H) and Simpson's index (D) were calculated by using PAST (Hammer *et al.*, 2001).

Nucleotide sequence accession numbers
The *g5* nucleotide sequences determined in this study have been deposited in GenBank under accession number of EU929030 to EU929055.

Results and discussion

Testing the g5 primers
The *g5* gene-specific primers (MCP-109F and MCP-368R) were tested against 22 *Roseobacter* and 4 *Rhodobacter* strains representing diverse Rhodobacterales isolates (Table 2). For all isolates shown in Table 2, PCR products of the expected size (782–794 bp) were recovered and verified to be *g5* by sequencing. All but one of a collection of *Roseobacter* and *Rhodobacter* strains isolated from Chesapeake Bay yielded a PCR product of the appropriate size with this primer set (Table 2; CB and AB, unpublished data). These results indicate that this primer set is suitable for recovering *g5* gene sequences from diverse groups of *Roseobacter* and *Rhodobacter*. Application of the *g5* primer set to four microbial communities collected from different locales in Chesapeake Bay yielded one specific amplicon of the expected size (data not shown), pointing to the suitability of this primer for the amplification of *g5* genes from environmental samples. To explore the diversity of *g5* genes and their spatial and temporal variations in the Chesapeake Bay, the PCR products from these four samples were cloned and sequenced.

Diverse and unique Roseobacter *in the Chesapeake Bay*
GTA diversity was assessed for four natural microbial assemblages representing upper, middle and lower bay stations (Supplementary Figure S1) during winter (March 2003) and middle bay in summer (July 2003). A total of 158 *g5* sequences were recovered from these four clone libraries. Phylogenetic analysis of the environmental *g5* clone sequences placed them all within *Roseobacter* and *Rhodobacter* clades and revealed 12 distinct phylogenetic clusters (designated as A–L; Figure 1). Eleven of these clusters fell within the *Roseobacter* clade and one cluster belonged to the *Rhodobacter* clade.

Among the 12 clusters retrieved from the Chesapeake Bay, only two clusters contain cultivated representatives (clusters E and J; Figure 1). Other clusters were divergent from known g5 sequences derived from either genome-sequenced Rhodobacterales strains or our bay isolates (<90% amino-acid identity). This finding mirrors recent 16S rRNA gene analysis-based studies of these same communities and further supports the finding that Chesapeake Bay contains unique *Roseobacter* clusters (Kan *et al.*, 2007, 2008). Additional efforts to isolate and characterize more indigenous roseobacters from the bay, particularly from winter samples, may help us better understand the distribution and ecology of these unique bacteria.

The coverage of each clone library ranged from 83% to 93% (Table 1). Rarefaction analysis showed that all clone libraries have a nearly sufficient number of clones to represent *g5* richness (Supple-

Table 2 Summary of *Roseobacter* and *Rhodobacter* strains used in this study

Strain designation	*Sampling site and date*	*Phylogenetic affiliation*[a]	*Reference*
CB1005[b]	Middle Chesapeake Bay, July 2007	Bacterium enrichment culture clone EB39.6, (98%) EU573108	This study
CB1006	Middle Chesapeake Bay, July 2007	*Silicibacter pomeroyi* DSS-3, (99%) AF098491	This study
CB1040	Middle Chesapeake Bay, July 2007	Uncultured Rhodobacteraceae bacterium clone DS153, (100%) DQ234235	This study
CB1051	Middle Chesapeake Bay, July 2007	Uncultured bacterium clone 2C229209, (100%) EU800950	This study
CB1024	Lower Chesapeake Bay, July 2007	Uncultured bacterium clone D23, (96%) AY375140	This study
CB1025	Lower Chesapeake Bay, July 2007	Uncultured bacterium clone NH10_46, (99%) DQ372853	This study
CB1049	Lower Chesapeake Bay, July 2007	Uncultured alphaproteobacterium clone 06-03-31, (97%) DQ153131	This study
CB1079	Upper Chesapeake Bay, July 2007	*Loktanella hongkongensis* strain UST950701-009W (99%), AY600301	This study
Sulfitobacter sp.VA-1	Virginia Beach, March 2008	*Sulfitobacter* sp. p66, (100%) EU864265	This study
Roseobacter sp.VA-2	Virginia Beach, March 2008	*Sulfitobacter dubius* strain KMM 3554T, (99%) AY180102	This study
Sulfitobacter sp. VA-6	Virginia Beach, March 2008	*Sulfitobacter* sp. p66, (99%) EU864265	This study
Sulfitobacter sp. VA-4	Virginia Beach, March 2008	*Sulfitobacter* sp. p66, (99%) EU864265	This study
Sulfitobacter sp. EE-36	Salt marsh, coast of Georgia	AF007254	González *et al.*, 1996
Silicibacter sp. TM1040	Marine dinoflagellate culture	AY332662	Miller and Belas, 2004
Roseovarius nubinhibens ISM	Caribbean Sea	AF098495	González *et al.*, 2003
Silicibacter pomeroyi DSS-3	Coastal Georgia, 1999	AF098491	González *et al.*, 2003; Moran *et al.*, 2004
CT-7	Chesapeake Bay, 2008	*Pseudorhodobacter incheonensis* strain KOPRI 13537, (99%) DQ001322	This study
CBB401	Chesapeake Bay Beach, April 2008	*Rhodobacter* sp. DQ12-45T, (96%) EF186075	This study
CBB404	Chesapeake Bay Beach, April 2008	Rhodobacteraceae bacterium D11-58, (96%) AM403233	This study
Rhodobacter capsulatus B10		M34129	Marrs, 1974

[a]NCBI (National Center for Biotechnology Information) accession number is provided for previously described strains. Closest match, sequence similarity and accession number (NCBI) is provided for strains isolated during this study.
[b]Six more strains closely related to this isolate were also tested.

mentary Figure S3). The diversity of *g5* gene sequences varied with sampling seasons and sites. In wintertime, both Shannon–Weaver and Simpson indices reveal that *g5* diversity is higher in the middle and upper bay compared with the lower bay (Tables 1 and 3). The summer sample had lower *g5* diversity compared with the winter sample (Tables 1 and 3).

Variation of GTA capsid genotypes along the Chesapeake Bay

A spatial variation of *g5* composition from upper to lower bay during the wintertime was evident (Table 3). Clusters A and B were unique to the upper bay, whereas clusters C and D occurred only in the lower bay. Clusters G–L were present in all three of the stations during winter with variable clone distribution frequency. Unlike the upper and lower bay, no unique cluster was found in the middle bay sample.

Cluster A constitutes more than 20% of the *g5* clones in the upper bay sample and is more closely related to *Rhodobacter* clade (Figure 1). A prior investigation of the phylogenetic diversity of the microbial community from this same sample revealed that members of the *Rhodobacter* group are abundant, representing $>15\%$ of the clones in a 16S rRNA gene library (Kan *et al.*, 2007). Several *Rhodobacter* strains have complete genome sequences available and are known to contain the GTA gene clusters (Lang and Beatty, 2007; Biers *et al.*, 2008). As more *Rhodobacter* genome sequences become available, it may be prudent to re-evaluate whether specific *g5* primer sets can be designed to discriminate between the *Rhodobacter* and *Roseobacter* groups.

Clusters C and D appeared only in the lower bay in winter and were most closely related to *Roseobacter* sp. CCS2 (88% and 89% amino-acid identity, respectively; Figure 1), which was isolated from Pacific coastal waters (http://www.roseobase.org/roseo/ccs2.html). The lower bay station is near the mouth of Chesapeake Bay, and is largely influenced by Atlantic coastal waters. Therefore, the lower bay has higher salinity (23 p.p.t.) than the upper and middle bay (Table 1). Thus, it is possible that these two clusters are more representative of typical coastal ocean roseobacters in comparison to other clusters recovered in this study.

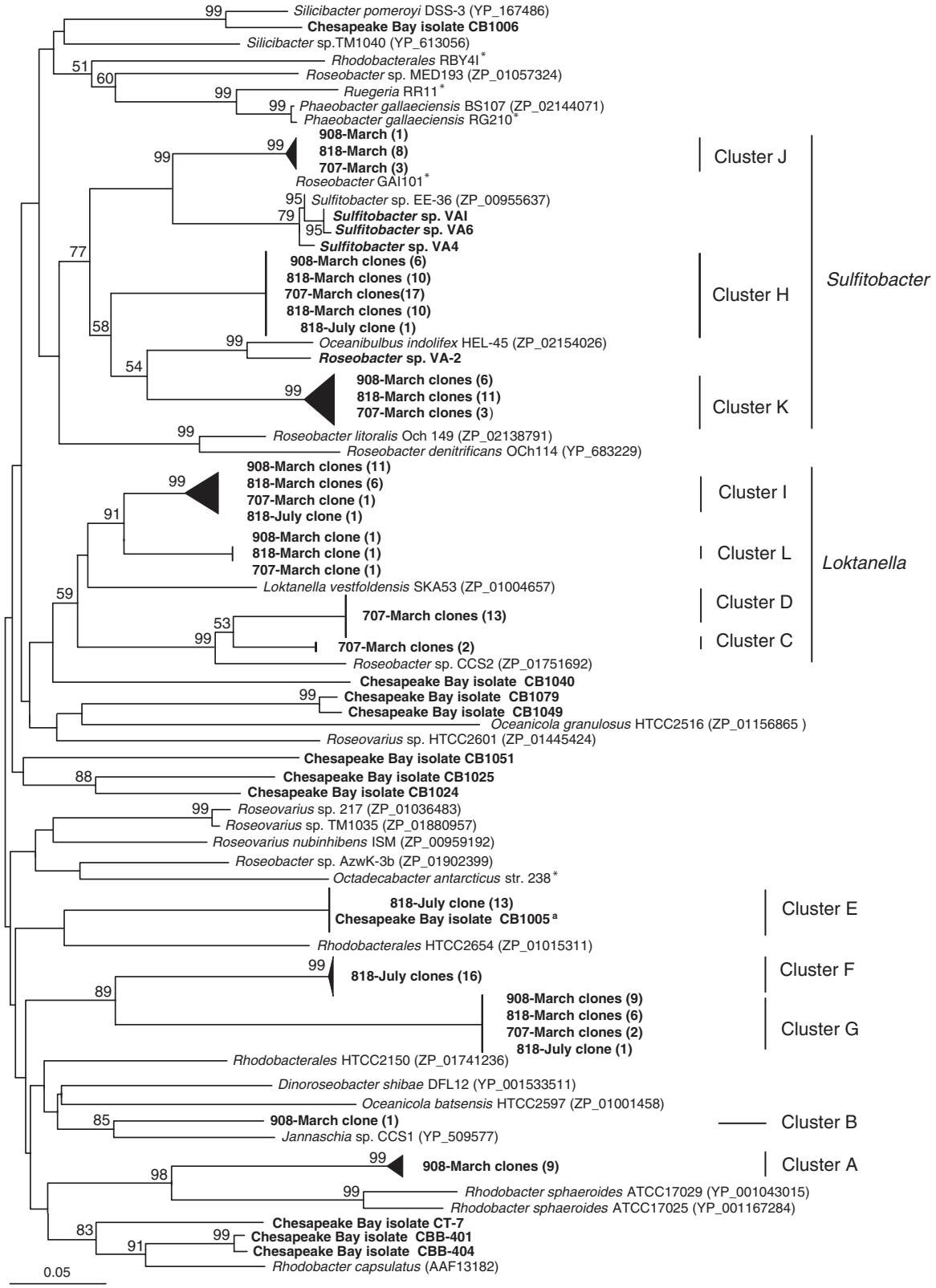

Figure 1 Neighbor-joining phylogenetic tree based on partial *g5* amino-acid sequences (ca. 250 aa) showing the phylogenetic diversity of *g*5 from Chesapeake Bay bacterial communities and *Roseobacter* and *Rhodobacter* isolates (bold type). Bootstrap numbers are shown as percentages based on 1000 replicates, and values of less than 50 were omitted. Numbers in the parentheses represent the number of closely related clones in that cluster *per se*. *Strains without accession number, but can be found at http://research.venterinstitute.org/moore/. [a]Six additional Chesapeake Bay isolates (CB1032, CB1023, CB1028, CB1083, CB1030 and CB1088) are closely related to CB1005.

Table 3 Comparison of *g5* composition and distribution in four clone libraries

Cluster	*Upper bay (station 908) March 2003*	*Middle bay (station 818) March 2003*	*Lower bay (station 707) March 2003*	*Middle bay (station 818) July 2003*
A	9 (20.5)[a]			
B	1 (2.3)			
C			2 (4.8)	
D			13 (31.0)	
E				13 (40.0)
F				16 (53.3)
G	9 (20.5)	6 (14.3)	2 (4.8)	1 (3.3)
H	6 (13.6)	10 (23.8)	17 (40.5)	1 (3.3)
I	11 (25.0)	6 (14.3)	1 (2.4)	1 (3.3)
J	1 (2.3)	8 (19.0)	3 (7.1)	
K	6 (13.6)	11 (26.1)	3 (7.1)	
L	1 (2.3)	1 (2.4)	1 (2.4)	

[a]The number of clones (and relative percentage) for a given cluster in the library.

Distinct GTA capsid genotypes are found between winter and summer

Clusters E and F (Figure 1) together accounted for 93% of the summer clone library and were not detected in any winter samples (Table 3), suggesting that there was a major shift in the *Roseobacter* population composition between winter and summer. This strong seasonal pattern for population composition was also evident in 16S rRNA gene clone libraries (Kan *et al.*, 2007, 2008). PCR amplicon yields of the *g5* gene from the summer samples were lower compared with the spring samples (data not shown). We therefore speculate that *Roseobacter* abundance could be lower in summer than winter. A quantitative PCR analysis based on the *g5* gene could be used to test this hypothesis in the future.

Cluster E contains 13 clones derived from the middle bay summer library (July 2003) as well as 7 additional bay strains that were also isolated several years later from the middle bay during summer (July 2007). This result is consistent with the recurring seasonal pattern of bacterioplankton in the bay (Kan *et al.*, 2006) and suggests these isolates represent a stable, abundant and unique summer-type *Roseobacter* in the Chesapeake Bay. These seven closely related *Roseobacter* isolates (represented by CB1005 in Table 2) also have no close match to publicly available clone or isolate sequences based on 16S rRNA gene homology searches (Table 2). Thus, characterization of these strains may reveal features that contribute to the success of these populations during the summer season. Another unique summer *Roseobacter* population (represented by 16 environmental clones) is cluster F (Figure 1). However, no Chesapeake Bay isolates fell within cluster F.

Most of the winter clone sequences are related to *Sulfitobacter* or *Loktanella* genera. Clusters J, H and K are affiliated with the *Sulfitobacter* group and likely represent the cold-adapted *Roseobacter* species (Figure 1). It has been demonstrated earlier that *Sulfitobacter*-related clones dominate the winter Chesapeake Bay bacterioplankton community (Kan *et al.*, 2007). Cluster J contains 12 clones recovered from all three winter libraries and strain *Roseobacter* sp. GAI-101 isolated from Georgia coastal sea water (USA) in winter (González *et al.*, 1999). Nearly all of the closest 16S rRNA gene sequences in GenBank to GAI-101 represent clones or isolates recovered from polar regions. Taken together, these data suggest that the unique winter clones in clusters J, H and K may represent roseobacters adapted to the winter cold environment in a temperate estuary. Several distinct *Roseobacter* clusters have been identified in temperate and polar regions based on the 16S rRNA gene marker (Selje *et al.*, 2004; Prabagaran *et al.*, 2007).

The *g5* clones in clusters G and I dominated the upper bay and their abundance declined from the upper to lower bay during the winter time. Such a trend could be related to the distribution pattern of phytoplankton *Chl* a (Table 1). Phytoplankton (that is, diatoms and dinoflagellates) blooms occur frequently during the winter–spring period in the upper Chesapeake Bay. It has been demonstrated that phytoplankton biomass and temperature are the two main factors regulating the bacterial population dynamics in the Bay (Kan *et al.*, 2006).

Congruency between the 16S rRNA gene and g5 *gene markers*

The comparison between *g5* and 16S rRNA gene phylogenies showed that the majority of environmental clones retrieved from the same winter samples based on the two different gene makers come from the same *Roseobacter* groups (*Sulfitobacter* and *Loktanella*; Figures 2a and b). In March 2003, clones that belong to *Sulfitobacter* and *Loktanella* groups together accounted for 85% of all the *Roseobacter g5* clones and 80% of the *Roseobacter* 16S rRNA gene clones. The GTA *g5* gene appears to be a conserved gene marker for Rhodobacterales. It is likely that the GTA gene clusters were acquired prior to the separation of the

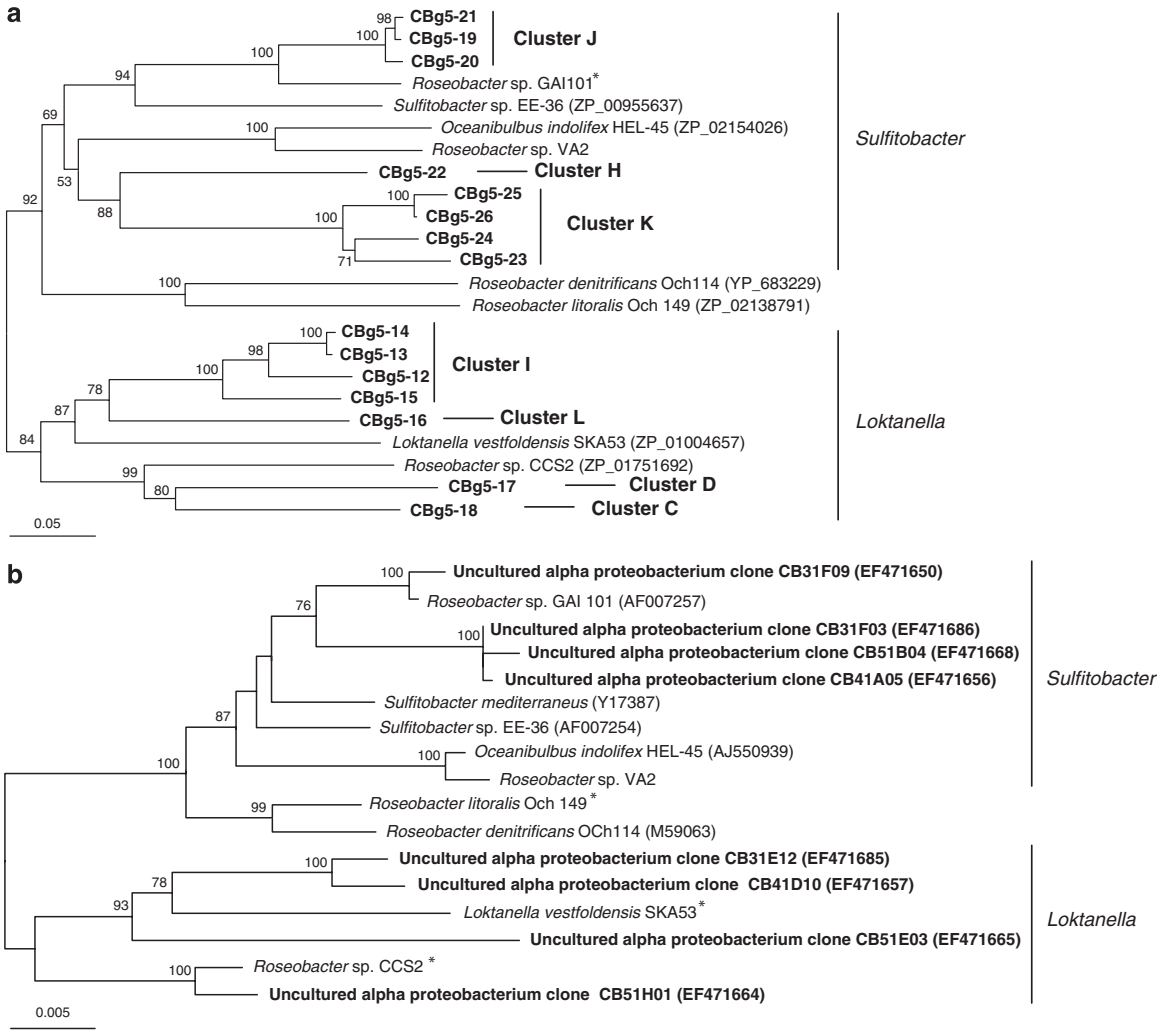

Figure 2 Phylogenetic trees based on *g5* nucleotide sequences (ca. 750 bp) (**a**) and 16S rRNA gene sequences (ca. 1300 bp) (**b**), respectively. Clones from Chesapeake Bay are shown in bold. Bootstrap numbers are shown as percentages based on 1000 replicate trees, and values of less than 50 were omitted. *Strains without accession number, but can be found at http://research.venterinstitute.org/moore/.

last common ancestor of all roseobacters and have been well preserved since then in different *Roseobacter/Rhodobacter* lineages with little lateral gene exchange (Lang and Beatty, 2007). Currently, no *g*5 genes from other bacterial groups fall into the Rhodobacterales group, more bacterial genome sequences will elucidate the monophylogeny of *g*5 for Rhodobacterales.

Exploring *Roseobacter* diversity and abundance based on the GTA capsid gene *g5* has several advantages: (1) all known *Roseobacter* genomes but one (strain HTCC2255), representing diverse members of *Roseobacter* clade, contain GTA gene clusters. The strain HTCC2255 seems to diverge from main stream roseobacters as it was isolated from oligotrophic waters and contains a much smaller genome (~2.3 Mb) compared with typical *Roseobacter* genomes (Biers *et al.*, 2008); (2) the GTA *g5* gene is highly conserved among the roseobacters; (3) each *Roseobacter* genome only contains a single copy of conserved *g5* gene and (4) the *g5* phylogeny is congruent with 16S rRNA gene phylogeny.

Conclusion

This is the first study examining the diversity of GTA capsid protein genes in the natural aquatic environment. The high diversity of *g5* sequences retrieved from Chesapeake Bay microbial communities and the concordance between *g5*- and 16S rRNA-based phylogenies demonstrate that the GTA capsid gene is an ideal group-specific gene marker for investigating the spatial and temporal distribution of *Roseobacter* and *Rhodobacter* groups in the natural environment. The high frequency of conservation of GTA genes in these abundant organisms suggests that GTA may be important in aquatic bacterial populations.

Acknowledgements

We acknowledge funding support from the National Science Foundation's Microbial Observatories Program (MCB-0132070, MCB-0238515, MCB-0537041 to FC), Biological Oceanography Program (OCE-0550485 to AB) and the funds from Xiamen University 111 program to FC and 2007CB815904 to NJ.

References

Alavi M, Miller T, Erlandson K, Schneider R, Belas R. (2001). Bacterial community associated with *Pfiesteria*-like dinoflagellate cultures. *Environ Microbiol* **3**: 380–396.

Biers EJ, Wang K, Pennington C, Belas R, Chen F, Moran MA. (2008). Occurrence and expression of gene transfer agent (GTA) genes in marine bacterioplankton. *Appl Environ Microbiol* **74**: 2933–2939.

Brinkhoff T, Giebel H, Simon M. (2008). Diversity, ecology, and genomics of the *Roseobacter* clade: a short overview. *Arch Microbiol* **189**: 531–539.

Brinkmeyer R, Knittel K, Jürgens J, Weyland H, Amann R, Helmke E. (2003). Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. *Appl Environ Microbiol* **69**: 6610–6619.

Brown MV, Bowman JP. (2001). A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiol Ecol* **35**: 267–275.

Buchan A, González JM, Moran MA. (2005). Overview of the marine *Roseobacter* lineage. *Appl Environ Microbiol* **71**: 5665–5677.

Budinoff CR, Hollibaugh JT. (2007). Ecophysiology of a Mono Lake picocyanobacterium. *Limnol Oceanogr* **52**: 2484–2496.

Choudhary M, Fu YX, Mackenzie C, Kaplan S. (2004). DNA sequence duplication in *Rhodobacter sphaeroides* 2.4.1: evidence of an ancient partnership between chromosomes I and II. *J Bacteriol* **186**: 2019–2027.

Choudhary M, Zanhua X, Fu YX, Kaplan S. (2007). Genome analyses of three strains of *Rhodobacter sphaeroides*: evidence of rapid evolution of chromosome II. *J Bacteriol* **189**: 1914–1921.

Crump BC, Armbrust EV, Baross JA. (1999). Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. *Appl Environ Microbiol* **65**: 3192–3204.

Crump BC, Hopkinson CS, Sogin ML, Hobbie JE. (2004). Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. *Appl Environ Microbiol* **70**: 1494–1505.

Giovannoni SJ. (1991). The polymerase chain reaction. In: Stackebrandt E, Goodfellow M (eds). *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons: New York, pp 177–201.

González JM, Covert JS, Whitman WB, Henriksen JR, Mayer F, Scharf B *et al.* (2003). *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., DMSP demethylating bacteria from marine environments. *Int J Syst Evol Microbiol* **53**: 1261–1269.

González JM, Kiene RP, Moran MA. (1999). Transformation of sulfur compounds by an abundant lineage of marine bacteria in the alpha-subclass of the class of *Proteobacteria*. *Appl Environ Microbiol* **65**: 3810–3819.

González JM, Simo R, Massana R, Covert JS, Casamayor EO, Pedoros-Alio C *et al.* (2000). Bacterial community structure associated with dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl Environ Microbiol* **66**: 4237–4246.

González JM, Whitman WB, Hodson RE, Moran MA. (1996). Identifying numerically abundant culturable bacteria from complex communities: an example from a lignin enrichment culture. *Appl Environ Microbiol* **62**: 4433–4440.

Hammer Ø, Harper DAT, Ryan PD. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electronica* **4**: 9. http://palaeo-electronica.org/2001_1/past/issue1_01.htm.

Hiraishi A, Ueda Y. (1995). Isolation and characterization of *Rhodovulum strictum* sp. nov. and some other purple nonsulfur bacteria from colored blooms in tidal and seawater pools. *Int J Syst Bacteriol* **45**: 319–326.

Kan J, Evans SE, Chen F, Suzuki MT. (2008). Novel estuarine bacterioplankton in rRNA operon libraries from the Chesapeake Bay. *Aquat Microb Ecol* **51**: 55–66.

Kan J, Suzuki MT, Wang K, Evans SE, Chen F. (2007). High temporal but low spatial heterogeneity of bacterioplankton in the Chesapeake Bay. *Appl Environ Microbiol* **73**: 6776–6789.

Kan J, Wang K, Chen F. (2006). Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE). *Aquat Microb Ecol* **42**: 7–18.

Lang AS, Beatty JT. (2000). Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci USA* **97**: 859–864.

Lang AS, Beatty JT. (2001). The gene transfer agent of *Rhodobacter capsulatus* and 'constitutive transduction' in prokaryotes. *Arch Microbiol* **175**: 241–249.

Lang AS, Beatty JT. (2002). A bacterial signal transduction system controls genetic exchange and motility. *J Bacteriol* **184**: 913–918.

Lang AS, Beatty JT. (2007). Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends Microbiol* **15**: 54–62.

Lilburn TG, Haith CE, Prince RC, Beatty JT. (1992). Pleiotropic effects of pufX gene deletion on the structure and function of the photosynthetic apparatus of *Rhodobacter capsulatus*. *Biochim Biophys Acta Bioenergetics* **1100**: 160–170.

Miller TR, Belas R. (2004). Dimethylsulfoniopropionate metabolism by *Pfiesteria*-associated *Roseobacter* spp. *Appl Environ Microbiol* **70**: 3383–3391.

Marrs B. (1974). Genetic recombination in *Rhodopseudomonas capsulata*. *Proc Nat Acad Sci USA* **71**: 971–973.

Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP *et al.* (2004). Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**: 910–913.

Nereng KS, Kaplan S. (1999). Genomic complexity among strains of the facultative photoheterotrophic bacterium *Rhodobacter sphaeroides*. *J Bacteriol* **181**: 1684–1688.

Paul JH. (2008). Prophages in marine bacteria*:* dangerous molecular time bombs or the key to survival in the seas? *ISME J* **2**: 579–589.

Prabagaran SR, Manorama R, Delille D, Shivaji S. (2007). Predominance of *Roseobacter, Sulfitobacter*, *Glaciecola* and *Psychrobacter* in seawater collected off Ushuaia, Argentina, sub-Antarctica. *FEMS Microbiol Ecol* **59**: 342–355.

Ravenschlag K, Sahm K, Pernthaler J, Amann R. (1999). High bacterial diversity in permanently cold marine sediments. *Appl Environ Microbiol* **65**: 3982–3989.

Scolnik PA, Haselkorn R. (1984). Activation of extra copies of genes coding for nitrogenase in *Rhodopseudomonas capsulata*. *Nature* **307**: 289–292.

Selje N, Simon M, Brinkhoff T. (2004). A newly discovered *Roseobacter* cluster in temperate and polar oceans. *Nature* **427**: 445–448.

Solioz M, Marrs B. (1977). The gene transfer agent of *Rhodopseudomonas capsulata*: purification and characterization of its nucleic acid. *Arch Biochem Biophys* **181**: 300–307.

Solioz M, Yen HC, Marris B. (1975). Release and uptake of gene transfer agent by *Rhodopseudomonas capsulata*. *J Bacteriol* **123**: 651–657.

Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Tolli JD, Sievert SM, Taylor CD. (2006). Unexpected diversity of bacteria capable of carbon monoxide oxidation in a coastal marine environment, and contribution of the *Roseobacter*-associated clade to total CO oxidation. *Appl Environ Microl* **72**: 1966–1973.

Wagner-Döbler I, Biebl H. (2006). Environmental biology of the marine *Roseobacter* lineage. *Annu Rev Microbiol* **60**: 255–280.

Wall JD, Braddock K. (1984). Mapping of *Rhodopseudomonas capsulata nif* genes. *J Bacteriol* **158**: 404–410.

Wall JD, Weaver PF, Gest H. (1975). Genetic transfer of nitrogenase-hydrogenase activity in *Rhodopseudomonas capsulata*. *Nature* **258**: 630–631.

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

Yen HC, Marrs B. (1976). Map of genes for carotenoid and bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata*. *J Bacteriol* **126**: 619–629.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)