RESEARCH ARTICLE

Water mass and depth determine the distribution and diversity of *Rhodobacterales* in an Arctic marine system

Yunyun Fu[1], Kimberley F. Keats[2], Richard B. Rivkin[2] & Andrew S. Lang[1]

[1]Department of Biology, Memorial University of Newfoundland, St. John's, NF, Canada; and [2]Department of Ocean Sciences, Ocean Sciences Centre, Memorial University of Newfoundland, St. John's, NF, Canada

Correspondence: Andrew S. Lang, Department of Biology, Memorial University of Newfoundland, St. John's, NL A1B 3X9, Canada. Tel.: +1 709 864 7517; fax: +1 709 864 3018; e-mail: aslang@mun.ca

Received 2 November 2012; revised 29 January 2013; accepted 30 January 2013. Final version published online 27 February 2013.

DOI: 10.1111/1574-6941.12085

Editor: Gary King

Keywords
Baffin Bay; gene transfer agent; biogeography; *Roseobacter*.

Abstract

Marine *Rhodobacterales* are recognized as a widespread, abundant, and metabolically versatile bacterial group in the world's oceans. They also show a nearly universal conservation of the genes for production of gene transfer agents (GTAs), virus-like particles that mediate genetic exchange between cells. It is not yet clear what factors determine the distribution of the various taxonomic subgroups of this order. To address this question, we analyzed the *Rhodobacterales* communities in 10 seawater samples from northern Baffin Bay collected during September 2008. A conserved gene from the GTA gene cluster was used to characterize the *Rhodobacterales* community structure. A total of 320 clones from 10 clone libraries were sequenced, and 22 operational taxonomic units representing putative species and 13 clusters representing putative genera were identified. A cluster related to *Octadecabacter* comprised 59% of total clones from the northern Baffin Bay. Phylogenetic analysis of the clones showed that the *Rhodobacterales* communities had distinct compositions in the different water masses that were sampled. A change in community structure related to depth was also observed. Therefore, in northern Baffin Bay where two ocean currents meet and mix, the *Rhodobacterales* community structures were primarily determined by water mass and depth.

Introduction

Marine heterotrophic picoplankton, such as bacteria, display biogeography patterns over global and local scales. Nutrients, temperature, and mortality are the main determinants in the global patterns of distribution of marine bacteria (Pommier *et al.*, 2007; Fuhrman *et al.*, 2008; Taniguchi & Hamasaki, 2008), while at the local scale, the primary factors controlling bacterial diversity and distribution are more complex and variable and include physical factors such as temperature, salinity (Bouvier & del Giorgio, 2002), depth (Brown *et al.*, 2009), and overall hydrography (Pinhassi *et al.*, 2003), as well as biotic factors such as nutrient availability, rates of mortality, and the phytoplankton community (González *et al.*, 2000). Some bacterial lineages are cosmopolitan, such as SAR11 (García-Martínez & Rodríguez-Valera, 2000; Morris *et al.*, 2002), but most taxa are more restricted (Pommier *et al.*, 2007; Ghiglione *et al.*, 2012). Although the biogeography of marine microorganisms is actively studied [e.g. (Galand *et al.*, 2010; Ghiglione *et al.*, 2012; Morris *et al.*, 2012)], both the deterministic factors as well as the associated environmental correlates are not fully understood (Nemergut *et al.*, 2011; Hanson *et al.*, 2012).

Bacteria in the order *Rhodobacterales*, especially the *Roseobacter* clade, are widespread and abundant in marine environments (Buchan *et al.*, 2005; Brinkhoff *et al.*, 2008; Giebel *et al.*, 2009). The characterized and cultured representatives from marine systems show a large diversity of physiological attributes that allow them to exploit their environment (Buchan *et al.*, 2005; Wagner-Döbler & Biebl, 2006; Newton *et al.*, 2010). These include aerobic anoxygenic photosynthesis (Shiba, 1991; Allgaier *et al.*, 2003), carbon monoxide oxidation (Moran *et al.*, 2007; Cunliffe, 2011), dimethylsulfoniopropionate (DMSP) metabolism and transformation (González *et al.*, 2000; Vila-Costa *et al.*, 2010), symbiosis with and pathogenesis of eukaryotes (Green *et al.*, 2004; Webster *et al.*, 2004; Seyedsayamdost *et al.*, 2011), degradation of hydrocarbons (Buchan & González, 2010), and secondary

FEMS MICROBIOLOGY ECOLOGY

metabolite production (Brinkhoff *et al.*, 2004; Martens *et al.*, 2007).

The *Rhodobacterales* additionally show a nearly universal presence of the gene cluster for producing gene transfer agents (GTAs) (Lang & Beatty, 2007; Biers *et al.*, 2008; Paul, 2008; Newton *et al.*, 2010; Lang *et al.*, 2012; McDaniel *et al.*, 2012). GTAs are bacteriophage-like particles that transfer cellular genomic DNA between cells (Stanton, 2007; Lang *et al.*, 2012). There is evidence that production of GTAs is widespread in this group of bacteria, with GTA-mediated gene transfer demonstrated for *Ruegeria pomeroyi* DSS-3 (Biers *et al.*, 2008) and release of GTA structural protein to the extracellular environment observed in multiple lineages within the order (Fu *et al.*, 2010). An additional three *Rhodobacterales* isolates, *Ruegeria mobilis* 45A6, *Roseovarius nubinhibens* ISM, and *Nitratireductor* sp. 44B9s, make GTAs capable of gene transfer to different genera of bacteria (McDaniel *et al.*, 2010, 2012). The abundance of *Rhodobacterales* in marine environments coupled with their high conservation of the GTA genes has contributed to the speculation that presence of GTAs in marine viral communities could explain the high proportion of cellular genes that are detected in viral metagenomic datasets (Kristensen *et al.*, 2010; Lang & Beatty, 2010). The possible role of *Rhodobacterales* in promoting genetic exchange in natural environments makes them even more compelling model microorganisms.

Despite the assumed ecological importance of *Rhodobacterales*, their distribution patterns and their response to environmental drivers are poorly understood. However, it is known that distinct distribution patterns exist. For example, the RCA cluster [also referred to as DC5-80-3 (Buchan *et al.*, 2005)] is abundant in polar oceans (Giebel *et al.*, 2009, 2011), but has not been reported from tropical/subtropical environments (Selje *et al.*, 2004). Although such distribution patterns are known for selected representatives of the *Roseobacter* clade, the patterns for most *Rhodobacterales* members have not been described. Several studies have used the 16S rRNA gene marker to track changes in the relative abundance of total *Rhodobacterales* communities in the South Atlantic basin (Morris *et al.*, 2012) and during a coastal six-year time series (Gilbert *et al.*, 2012), but such an approach fails to capture the finer detail expected within this physiologically diverse order. The *g5* gene in GTA gene cluster encodes the major capsid protein of GTA particles. It has been used as a higher resolution phylogenetic marker to document the seasonal and spatial patterns in the structure of *Rhodobacterales* communities in both temperate and sub-Arctic locations (Zhao *et al.*, 2009; Fu *et al.*, 2010). This sequence generally produces a similar phylogeny with 16S rRNA gene (Lang & Beatty, 2007; Biers *et al.*, 2008; Zhao *et al.*, 2009) but provides a more detailed view of *Rhodobacterales* diversity (Zhao *et al.*, 2009; Fu *et al.*, 2010).

In this study, we used *g5* sequences to examine *Rhodobacterales* communities in northern Baffin Bay in the Canadian Arctic. During the observation period, this region was characterized by distinct water masses where colder and lower salinity Arctic water merged with the warmer and higher salinity water of the western Greenland Current (Melling *et al.*, 2001; Motard-Côté *et al.*, 2012). These water masses have distinct physicochemical and biological characteristics, and we examined whether the diversity and distribution patterns of *Rhodobacterales* communities differed as a function of water mass characteristics and environmental parameters.

Materials and methods

Sample sites, measurements of physicochemical parameters and bacterial community characteristics

Samples were collected during the Surface Ocean-Lower Atmosphere Study (SOLAS) expedition on the CCGS *Amundsen* in September 2008. A total of 17 stations were sampled (Motard-Côté *et al.*, 2012), among which 10 samples from five stations were included in this study (Fig. 1, Table 1). Samples were analyzed from the depths of the 50% surface light intensity and from the depth of the maximum chlorophyll *a* (chl *a*) concentration at stations 111, 126, 141, and 202. An additional deep sample was analyzed from 71 m at station 111, and one sample was analyzed from station 134 from the depth of 50% surface light intensity. Water samples were collected using a rosette sampler equipped with 12-L Niskin bottles.

Vertical profiles of physicochemical, biological, and hydrographical parameters including depth, temperature, *in vivo* chlorophyll fluorescence, salinity, water density, dissolved O_2 concentration, pH, nitrate concentration, and photosynthetic active radiance (PAR) were measured at each sampling station by a Seabird 911*plus* conductivity–temperature–depth (CTD) probe, equipped with a fluorometer (SeaPoint 2465) and an irradiance sensor (QCP-2300, Biospherical Instruments), mounted on the rosette. *In situ* chlorophyll fluorescence was calibrated against measurements of extracted chl *a* from water samples collected at five depths between 5 and 100 m, following the acidification method (Parsons *et al.*, 1984). Extracted chl *a* concentrations were determined with a Turner Designs Model 10-AU fluorometer.

Heterotrophic picoplankton abundance (HPA) was determined by flow cytometry using a FACSort™ (Becton Dickinson) equipped with a 488-nm argon-ion laser

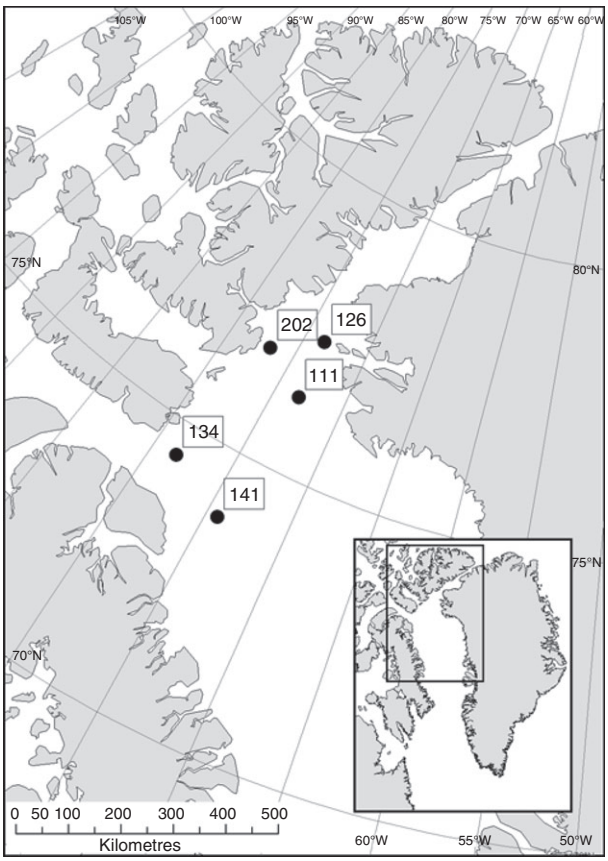

Fig. 1. Sampling stations for this study, which were part of the Arctic Surface Ocean-Lower Atmosphere Study (SOLAS) during the CCGS *Amundsen* cruise in September 2008.

(Li, 1995). Samples for flow cytometry were fixed in a final concentration of 1% paraformaldehyde for 10 min at room temperature, rapidly frozen in liquid nitrogen within 1 h of sample collection, and stored at −80 °C until analysis. Samples were stained with SYBR Green 1 and analyzed using standard protocols (Marie *et al.*, 1999; Li & Dickie, 2001). Flow cytometry counts were verified by comparison with acridine orange direct counts (Hobbie *et al.*, 1977). Bacterial production (BP) was calculated from [^{14}C]-leucine incorporation rates during 6-h dark incubations using the final concentration of 10 nM leucine and a standard conversation factor of 3.1 kg C mol^{-1} [^{14}C]-leucine incorporated (Simon & Azam, 1989). Bacterial biomass (BB) was estimated from bacterial abundance and cellular carbon content, computed from the geometric mean cell volume of each sample (Hale *et al.*, 2006). The bacterial growth rate (SGR) was calculated assuming exponential growth of bacteria during the bacterial production incubations: SGR = $\log_e[(BP_{\Delta}t + BB)/BB]/\Delta t$, where $BP_{\Delta t}$ is bacterial production obtained during the incubation and Δt is the incubation time in days. The average volumetric HPA, BP, and SGR were calculated for the surface mixed layer (from the bottom of the mixed layer to the pycnocline) and the deep layer (below the pycnocline to 200 meters) at the included sampling stations.

DNA extractions and PCR amplification, cloning, and sequencing of GTA *g5* genes

Water samples (500 mL) were filtered onto 0.22-μm polycarbonate filters (Millipore), which were then stored at −20 °C until DNA was extracted. Total DNA was extracted from individual filters in 2-mL lysis buffer (10 mM Tris-HCl buffer with 1 mM EDTA and 0.8 M sucrose, pH 8) with 1 μg mL^{-1} lysozyme at 37 °C for 30 min. This was followed by addition of 200 μL of 10% (w/v) cetyltrimethylammonium bromide (CTAB) in 1.4 M NaCl and incubation at 65 °C for 30 min. Samples were sequentially extracted with phenol–chloroform–isoamyl alcohol (25 : 24 : 1) and chloroform–isoamyl alcohol (24 : 1), followed by isopropanol precipitation of DNA. The DNA pellets were washed with 70% ethanol, air-dried, and dissolved in 200-μL Tris-EDTA buffer (pH 8).

The *g5* gene fragments (approximately 780 bp) were amplified from the extracted DNA using primers MCP-3F (5′-GGCTAYCTGGTSGATCCSCARAC-3′) and MCP-4R (5′-TAGAACAGSACRTGSGGYTTKGC-3′) described previously (Zhao *et al.*, 2009). The target gene fragments were amplified by PCR in 25 μL volumes containing 0.5 μM of each primer, 3% DMSO (v/v), 200 μM dNTPs, 0.5 U Phusion Hot Start polymerase (Finnzymes), and 1× Phusion GC buffer (Finnzymes). Thermocycling conditions were as follows: 30 s at 98 °C followed by 30 cycles of 98 °C for 10 s, 60 °C for 10 s, and 72 °C for 30 s, and a final incubation at 72 °C for 5 min. The reactions were then incubated with 1 U of *Taq* polymerase (NEB) at 72 °C for 10 min in 1× *Taq* buffer (NEB) to add 3′ adenine overhangs, and PCR products were purified using a MinElute PCR purification kit (Qiagen). Amplicons were cloned into pGEM-T Easy (Promega). The clones containing appropriately sized inserts were sequenced at the Center for Applied Genomics (Toronto, Canada) using the M13 primers that bind to the cloning vector.

Sequence and phylogenetic analyses

The *g5* sequences from this study have been deposited in the NCBI GenBank database, accession numbers KC111451–KC111767. After removal of the primer sequences, the *g5* nucleotide sequences were aligned according to the corresponding aligned amino acid sequences with ClustalW

Table 1. Physicochemical and biological characteristics at the stations studied

Station	Latitude, longitude	Sampling date	Depth (m)	Sample feature*	Temperature (°C)	Chl *a*† (μg L^{-1})	Salinity (psu)	O_2 concentration (mL L^{-1})	PAR‡ (quanta m^{-2} s^{-1})	HPA§ (cells L^{-1})	BP¶ (μg C L^{-1} day^{-1})	SGR** (day^{-1})	Water mass classification††
202	76°48′ N, 76°48′ W	16-9-2008	6	50% L	−0.96	0.09	28.36	8.61	7.14×10^{18}	3.79×10^{8}	2.43	0.29	ASW
			35	chl max	−0.36	0.15	30.2	8.67	0.14×10^{18}	3.11×10^{8}	1.50	0.19	ASW
134	74°25′ N, 78°25′ W	9-9-2008	6	50% L	3.31	0.07	32.13	7.55	23.65×10^{18}	16.6×10^{8}	0.36	0.01	BBSW
111	76°18′ N, 73°14′ W	13-9-2008	8	50% L	3.81	0.13	31.51	7.61	11.87×10^{18}	4.23×10^{8}	0.84	0.09	BBSW
			26	chl max	3.34	0.46	32.02	7.91	2.19×10^{18}	5.41×10^{8}	0.58	0.06	BBSW
			71	0.2% L	−0.85	0.14	33.4	7.45	0.07×10^{18}	2.72×10^{8}	0.20	0.04	NA
126	77°20′ N, 73°25′ W	18-9-2008	6	50% L	3.41	0.13	31.52	7.58	55.25×10^{18}	9.97×10^{8}	1.10	0.05	BBSW
			14	chl max	1.17	0.26	32.08	7.85	17.01×10^{18}	8.44×10^{8}	1.03	0.07	BBSW
141	73°52′ N, 74°17′ W	23-9-2008	7	50% L	1.3	0.11	31.55	7.76	8.39×10^{18}	5.38×10^{8}	0.68	0.06	INT
			38	chl max	1.02	0.17	32.34	7.75	0.45×10^{18}	4.93×10^{8}	0.34	0.03	INT

*50% L: 50% surface light intensity; chl max: highest chlorophyll *a* concentration in the water column; 0.2% L: 0.2% surface light intensity.
†Chlorophyll *a* concentration.
‡Photosynthesis active radiation.
§Heterotrophic picoplankton abundance.
¶Bacterial production.
**Specific growth rate.
††ASW, Arctic surface water; BBSW, Baffin Bay surface water; INT, intermediate water mass; NA, sample was excluded from water mass segregation.

(Thompson *et al.*, 1994) within MEGA 4 (Tamura *et al.*, 2007). The evolutionary divergence was calculated based on the nucleotide alignments. Sequences with $\geq$ 95% identity were assigned to operational taxonomic units (OTUs). This criterion represented a species-level distance as inferred by comparing the currently available *g5* sequences from the genome-sequenced *Rhodobacterales* representatives. The coverage (C) of each clone library was estimated using the formula $C = 1-(N/n)$, where N is the number of unique sequences in the sample and n is the total number of sequences in the library (Good, 1953). Rarefaction analysis was conducted using Analytical Rarefaction v1.3 (http://strata.uga.edu/software/index.html) as another estimate of the proportion of the population diversity sampled. The evolutionary distances of *g5* sequences were calculated based on the alignment using the maximum composite likelihood method (Tamura *et al.*, 2004), and the phylogenetic trees were constructed by the neighbor-joining method using the pairwise deletion of gaps option (Saitou & Nei, 1987).

Community composition comparisons

The *Rhodobacterales* sequences were grouped at the species level using $\geq$ 95% nucleotide identity and at the genus level using $\geq$ 80% identity criteria for the *g5* sequences. These values were inferred by comparing the *g5* sequences from available *Rhodobacterales* genome sequences. The grouping of OTUs and clusters was subsequently revised according to the phylogenetic analysis. The community compositions were compared at both the individual sample and water mass scales using UniFrac (Lozupone & Knight, 2005) based on the phylogenetic trees of the *g5* clones. The communities were clustered using the UniFrac jackknife analysis, and the robustness of nodes was estimated with 1000 iterations.

The Shannon index (Shannon & Weaver, 1949) was calculated to quantify the *Rhodobacterales* community diversity with the 95% nucleotide identity as the criterion for demarcation of species/OTUs. The diversity of *Rhodobacterales* communities in different water masses was compared by analysis of variance (ANOVA), with Tukey's *post hoc* tests to identify significant differences in diversity in the water masses. The relationships between *Rhodobacterales* diversity and environmental parameters were assessed by the Pearson correlation coefficient.

The relative distribution of a cluster or OTU was estimated by its percent proportional occurrence, $N/n \times 100$, where N is the number of clones in the OTU or cluster and n is the total number of clones sequenced from the sample. The relationships between the proportional occurrences for abundant phylotypes (defined as a cluster or OTU, which accounted for > 5% of total clones) and physicochemical parameters were examined using the Pearson correlation coefficient.

Results

Field site and physicochemical characteristics

The complete description for all 17 sampling stations in this SOLAS expedition is reported elsewhere (Motard-Côté *et al.*, 2012). Water temperatures ranged from − 0.96 to 3.81 °C for the samples analyzed in this study, with the coldest waters at station 202 and the warmest at station 111 (Fig. 1, Table 1). Salinity ranged from 28.35 to 33.40 psu (Table 1). A temperature–salinity plot (Supporting Information, Figure S1) showed the segregation of the SOLAS sampling stations into three water masses, which were classified as Arctic surface water (ASW), Baffin Bay surface water (BBSW), and an intermediate zone (INT) between the distinct BBSW and ASW masses. Lower salinity accompanied with lower temperatures was the defining feature of the ASW. Detailed characterization of the *Rhodobacterales* was performed on samples collected in all three water masses (Table 1), where station 202 represents ASW, stations 111, 126, and 134 represent BBSW, and station 141 represents INT. Based on the temperature–salinity characteristics above the depth of the pycnocline, the three water masses were distinct and were each hydrographically coherent. Below the depth of the pycnocline, there were no significant differences in the temperature–salinity relationships, suggesting that segregation of these water masses is restricted to the surface layer and depths above the pycnocline. Therefore, the sole deep sample, collected at 71 m from station 111, was excluded from analyses comparing the different water masses but it is still discussed in other contexts below.

The ASW had a higher O_2 concentration than BBSW and INT, but the differences in chl *a* concentration among water masses were not significant (ANOVA, $F = 0.4343$, $P = 0.667$). The sea surface PAR varied with stations, but also did not show significant inter-water mass variation. The HPA ranged from 3.11 to 16.64×10^8 cells L^{-1}, with the highest abundance in the BBSW and the lowest in the ASW. The BP showed the reverse pattern, where ASW had significantly higher BP than BBSW and INT (ANOVA, $F = 8.84$, $P = 0.016$, d.f. = 8; Tukey, $P < 0.05$). Consistent with BP, ASW also showed a higher SGR (ANOVA, $F = 18.28$, $P = 0.003$, d.f. = 8; Tukey, $P < 0.01$).

Description of GTA *g5* clone libraries

We created 10 clone libraries of GTA *g5* amplicons representing the 10 samples collected at the five stations with a

total of 320 *g5* clones being sequenced from the 10 libraries. All libraries were sequenced to $\geq$ 90% coverage (Table 2), although all the rarefaction curves did not reach saturation (Figure S2) indicating there was undetected diversity for some of the samples. Twenty-two OTUs were identified with the criterion of 95% nucleotide sequence identity, representing 22 putative species, and 13 different clusters were found with the 80% identity criterion, representing 13 putative genera. Of the 22 OTUs, 12 (54.5%) were singletons, which were only detected once in all samples. These singletons represented only 3.4% of all clones. The dominant OTU (i) accounted for 48.4% of all clones and was detected in all 10 samples. At the putative genus level, five singletons were found, representing 1.5% of all clones. Cluster A, mainly comprised of OTU i, was the most abundant cluster in all water masses.

The phylogenetic relationships of the clones from this study with reference strains are shown in Fig. 2. The clone sequences are very diverse within the *Rhodobacterales*. Some of the OTUs and clusters have well-supported close relationships to cultured strains with available *g5* sequences, but most do not (Fig. 2).

Assigning taxonomic labels to the environmental *Rhodobacterales* OTUs and clusters

Representative GTA *g5* sequences from known isolates that are available in the public sequence databases can be affiliated with only two of the 22 the OTUs (ii and vii) and three of the 13 clusters (A, B, and G) identified in the clone libraries (Table 4; Fig. 2). At the species level, *Oceanibulbus indolifex* HEL-45 clearly grouped with OTU ii and *Sulfitobacter* sp. EE-36 and NAS-14.1 grouped with OTU vii. The *Sulfitobacter* sequences formed a rare group (2%) of the detected sequences, and they were not detected in the ASW. The three samples where these clones were detected were all from the surface mixed layer at stations 134, 126, and 141 where water temperatures were relatively high (Table 1).

At the putative genus level, cluster A, which was the most abundant group of sequences in all three water masses (Fig. 4), is distantly related to the genus *Octadecabacter* (Fig. 2). Cluster A accounted for 91%, 53%, and 33% clones in the ASW, INT, and BBSW, respectively.

A number of other OTUs and clusters appeared affiliated with known *Loktanella* representatives, but none of these relationships were within our criteria for species or genera relationships or supported by bootstrapping of the phylogenetic tree (Fig. 2).

Correlation of *Rhodobacterales* diversity with environmental parameters

We calculated the Pearson correlation coefficients between the *Rhodobacterales* diversity (as estimated by Shannon index) and the physicochemical and biological parameters measured for the 10 samples (Table 3). The HPA and temperature showed significant positive correlations with Shannon index, whereas O_2 concentration and SGR showed negative correlations with Shannon index.

Comparison of the *Rhodobacterales* communities in different water masses

Rhodobacterales community characteristics, including diversity and community composition of the three water masses, were compared. The Shannon indices were significantly different among water masses (ANOVA,

Table 2. Characteristics of the 10 *g5* clone libraries

Sampling station	Water mass*	Sample feature†	Number of clone sequences	Number of OTUs	Number of clusters	Coverage (%)	Shannon index (*H'*)	OTUs in each sample
202	ASW	50% L	34	5	4	91	0.5547	i, iii, xxix, xxx, xxxi
202	ASW	chl max	36	3	3	94	0.2253	i, iii, x
134	BBSW	50% L	26	7	6	96	1.6852	i, ii, v, vi, vii, viii, xi
111	BBSW	50% L	48	8	6	90	0.9695	i, ii, iii, v, vi, x, xii, xiii
111	BBSW	chl max	26	6	5	100	1.6179	i, ii, iii, iv, v, vi
111	NA	0.2% L	17	3	3	100	0.8351	i, iv, v
126	BBSW	50% L	47	9	7	94	1.7781	i, ii, v, vi, vii, ix, xiv, xv, xvi
126	BBSW	chl max	33	8	6	91	1.5037	i, ii, iii, iv, v, vi, ix, x
141	INT	50% L	30	7	7	90	1.3911	i, ii, iii, iv, vii, viii, xvii
141	INT	chl max	23	4	2	91	0.8934	i, iii, xviii, xix

*ASW, Arctic surface water; BBSW, Baffin Bay surface water; INT, intermediate water mass; NA, sample was excluded from water mass segregation.

†50% L: 50% surface light intensity; chl max: highest chlorophyll *a* concentration in the water column; 0.2% L: 0.2% surface light intensity.

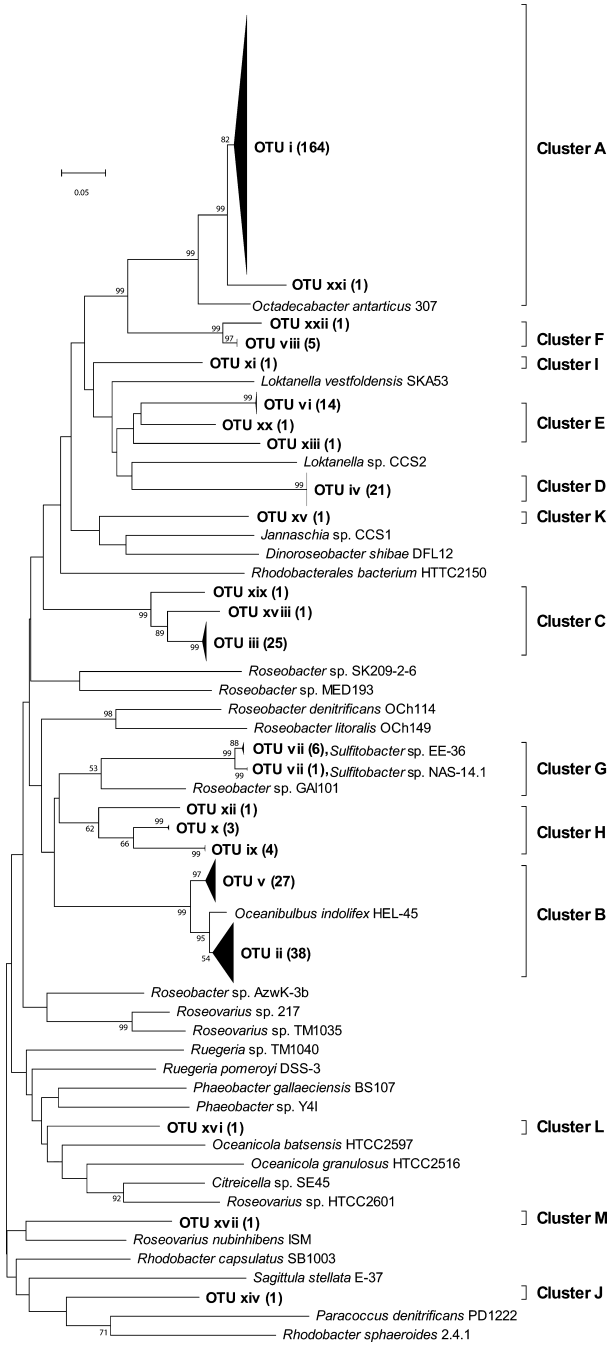

Fig. 2. Phylogenetic relationships of GTA *g5* sequences. The phylogenetic tree presents clone sequences detected and reference *g5* sequences from available genomes in the NCBI database. Branches for OTUs were collapsed, and the numbers in parentheses report the number of clones detected for each OTU. The assigned clusters are indicated on the right. The unrooted neighbor-joining tree is based on the aligned nucleotide sequences corresponding to the aligned translated amino acids sequences. Bootstrap values (percentage based on 1000 iterations) are shown for the branch points supported > 50%. The scale bar indicates the number of substitutions per site.

Table 3. Pearson correlation coefficients of environmental parameters with the Shannon index from the 10 samples and the proportional occurrence of the three most abundant sequence clusters

		Cluster*		
Parameter	Shannon index	A	B	C
Depth	−0.4230	−0.1630	−0.2857	0.0526
Temperature	**0.7387**†	−0.4009	0.6052	−0.0664
Chl *a*	0.2917	−0.3183	0.2362	0.1499
Salinity	0.4731	**−0.7664**	0.4531	0.1137
O_2	**−0.6624**	**0.7789**	−0.5701	−0.0275
Photosynthesis active radiation	0.6285	−0.3823	**0.7007**	−0.3859
Heterotrophic picoplankton abundance	**0.7112**	−0.4787	**0.7251**	−0.2508
Bacterial production	−0.4470	**0.6816**	−0.3124	−0.2681
Specific growth rate	**−0.6811**	**0.7893**	−0.5559	−0.1980

*Clusters are defined in Fig. 2.

†Numbers in bold font indicate significant correlations ($P < 0.05$).

P-value = 0.0148, *F*-ratio = 9.23, d.f. = 8) and were highest in the BBSW and lowest in the ASW, and the differences were significant (Tukey, $P = 0.012$). The INT diversity showed no significant difference (Tukey, $P > 0.05$) with either of the other two water masses (Fig. 3). However, the Shannon index quantifies the degree of complexity and does not indicate differences in composition or community structure. We therefore used UniFrac analyses (Lozupone & Knight, 2005) to compare the community compositions based on phylogenetic relationships of the *g5* sequences. This showed that samples from different water masses form separate clusters, with exception of the surface sample from station 111 (Fig. 4). Therefore, the different water masses show distinct *Rhodobacterales* community compositions. The differences are the most distinct and robust between ASW and BBSW samples, and INT samples were more related to ASW samples.

Different *Rhodobacterales* phylotypes showed different distribution patterns (Figs 4 and 5). Only three clusters (A, C, and F) were found in all three water masses, and only cluster A was found in all samples. Pearson correlation coefficient analysis was used to investigate the relationship between the proportional occurrence of the three most abundant clusters and the environmental parameters (Table 3). Cluster A showed significant positive correlation to O_2 concentration, BP, and SGR and a negative correlation to salinity. Cluster B showed significant positive correlation to PAR and HPA. Cluster C did not show significant correlations to any of the measured parameters.

Influence of depth on *Rhodobacterales* communities

Three samples were collected at station 111 in BBSW, and these showed that there is stratification of *Rhodobacterales* in the water column (Fig. 5). The chl *a* maximum depth had the highest diversity (Table 2). Clusters A and D had obvious and opposite differences in proportional occurrence according to depth (Fig. 5). Cluster A was mainly found in the surface layer and comprised 73% of the clones from this layer, while cluster D was more abundant in the deep samples. In contrast, cluster B was fairly evenly distributed throughout the water column.

Discussion

We characterized the diversity of bacteria in the order *Rhodobacterales* in northern Baffin Bay in the eastern Canadian Arctic. The diversity of these bacteria in this location is lower than previously reported in Newfoundland coastal water (Fu *et al.*, 2010) and in the Chesapeake Bay (Zhao *et al.*, 2009). This is consistent with the global trend of decreasing biodiversity from the equator to the poles (Martiny *et al.*, 2006). Furthermore, the composition of the *Rhodobacterales* community in northern Baffin Bay differs compared to temperate locations. For example, the dominant cluster we identified, which is related to *Octadecabacter*, was not found in the Chesapeake Bay. The *Rhodobacterales* sequences recovered in this study affiliate with sequences from the *Roseobacter* clade, as opposed to *Rhodobacter,* consistent with their previously recognized abundance in marine systems based on 16S rRNA gene (González & Moran, 1997; Buchan *et al.*, 2005; Gilbert *et al.*, 2012). Overall, however, identifiable isolates closely affiliated with the taxa we have identified here are few, and more efforts are needed to characterize cultured representatives to further explain *Rhodobacterales* biogeography. Quantification of specific bacterial groups using fluorescence *in situ* hybridization during this SOLAS expedition (Motard-Côté *et al.*, 2012) indicated that Roseobacters accounted for 12% and 6% of prokaryotes in the ASW and BBSW masses, respectively. However, these percentages should be treated with caution as current tools are likely unable to accurately quantify the *Roseobacter* clade as a whole (Giebel *et al.*, 2011), and subgroup-specific probes for lineages within the clade appear more reliable (Buchan *et al.*, 2009; Giebel *et al.*, 2009; Yao *et al.*, 2011) with the remaining problem that they do not provide data for the entire group.

We found water mass to be the most important factor for the distribution of marine *Rhodobacterales* communities in northern Baffin Bay, which agrees with previous studies looking at different levels of bacterial community composition (Baldwin *et al.*, 2005; Pommier *et al.*, 2005; Giebel *et al.*, 2009, 2011; Schattenhofer *et al.*, 2009). There are some general patterns observed for the distributions of the three predominant clusters in the different water masses (Fig. 4). Cluster A was the only group found in all samples in all water masses. Cluster B was not found in the ASW mass. Cluster C was found in all three water masses, but had its highest proportional occurrence in the INT samples. A comparison of the diversity among the three water masses in northern Baffin Bay shows that low diversity in ASW corresponded with high bacterial growth rate, as indicated by SGR, and the situation was reversed in BBSW. Overall diversity showed an inverse correlation with SGR, whereas the abundance of cluster A showed a strong positive correlation with SGR. This presumably reflects the dominance of this group, particularly in conditions favoring higher growth rates. In contrast, cluster B is correlated with light intensity and HPA, whereas cluster C did not show any significant correlations. Therefore, although some of the taxa distribution patterns correlate with some environmental factors, these environmental factors do not sufficiently explain all of the observed patterns. Marine *Rhodobacterales* have been shown to show a strong seasonal pattern in abundance at a specific temperate coastal location with peaks in abundance in concert with primary production (Gilbert *et al.*, 2012), which again reflects a combination of environmental factors. We do not think the presence of only one station from each of the ASW and INT in our analyses has artificially caused the observed water mass differences in diversity (Fig. 3) because the 50%

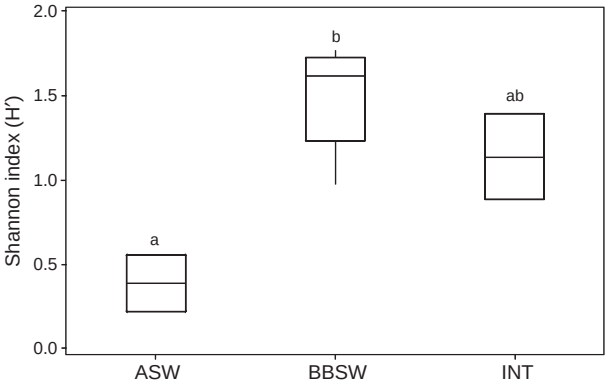

Fig. 3. Comparison of *Rhodobacterales* diversity in different water masses as quantified by Shannon index. The *Rhodobacterales* diversity was significantly lower in ASW (a) than BBSW (b) but there was no significant difference between the diversities of INT (ab) and ASW or BBSW. The horizontal line in the box is the mean Shannon index, and the bottom and top edges of the box are 25% and 75% quartiles, respectively. The vertical bars are the 95% confidence interval, when outside the quartiles demarcated by the box.

light and chl *a* maximum samples group together for the stations within the INT and ASW, while the same is not seen for the BBSW where there are three stations (Fig. 4). Using different criteria to assign all 17 SOLAS stations to either ASW or BBSW (Motard-Côté *et al.*, 2012), our INT station fell within the ASW, and the UniFrac community analysis does cluster our INT and ASW stations together (Fig. 4). However, the Shannon index measure of overall diversity (Fig. 3) supports the INT designation we decided upon from the temperature–salinity profiles (Figure S1). It would be ideal to have more samples from the INT and ASW masses, but samples for *Rhodobacterales* characterizations were only collected at a subset of the 17 SOLAS stations.

Depth affects the *Rhodobacterales* community structure, as is generally observed in studies of bacterial communities (Baltar *et al.*, 2007; Schattenhofer *et al.*, 2009; Morris *et al.*, 2012). Although it was not a statistically supported correlation (Table 3), perhaps because there were generally only two depths sampled at most of the stations, the depth profile from the three samples collected at station 111 shows distinct distributions for different sequence clusters (Fig. 5). The additional deep sample from this station came from below the pycnocline. Although the pycnocline is usually considered as a physical barrier of vertical mixing, we find it is not likely acting as a physical barrier in *Rhodobacterales* community distribution because all the taxa found in the 71-m sample are also found at the chl *a* maximum depth and the UniFrac analysis clustered these two samples (Fig. 4). One possibility for these bacteria to cross the pycnocline mixing barrier is by adhering to sinking partials, such as marine snow, and *Rhodobacterales* are known to form aggregates and adhere to particles (Gram *et al.*, 2002; Dang *et al.*, 2008). When reaching the deeper location, the environmental factors could further tune the community structure based on the physiological properties of each group. Sequences representing cluster D contributed 65% of clones to the deep sample but were only recovered in one surface sample, and there represented only 3% of the clones (Fig. 4), suggesting this group is adapted to deeper water conditions.

The dominant *Rhodobacterales* sequences found in this study are related to *Octadecabacter antarcticus* (based on > 80% nucleotide identity of *g5* fragments for genus-level relationships). This group was previously found to be abundant in polar oceans and sea ice (Staley & Gosink, 1999; Brinkmeyer *et al.*, 2003). All cultured *Octadecabacter* representatives have been obtained from polar environments and display low optimal growth temperatures of 4–15 °C (Gosink *et al.*, 1997). In our study, the occurrence of the *Octadecabacter*-related cluster is positively correlated to the SGR, which suggests that it is one

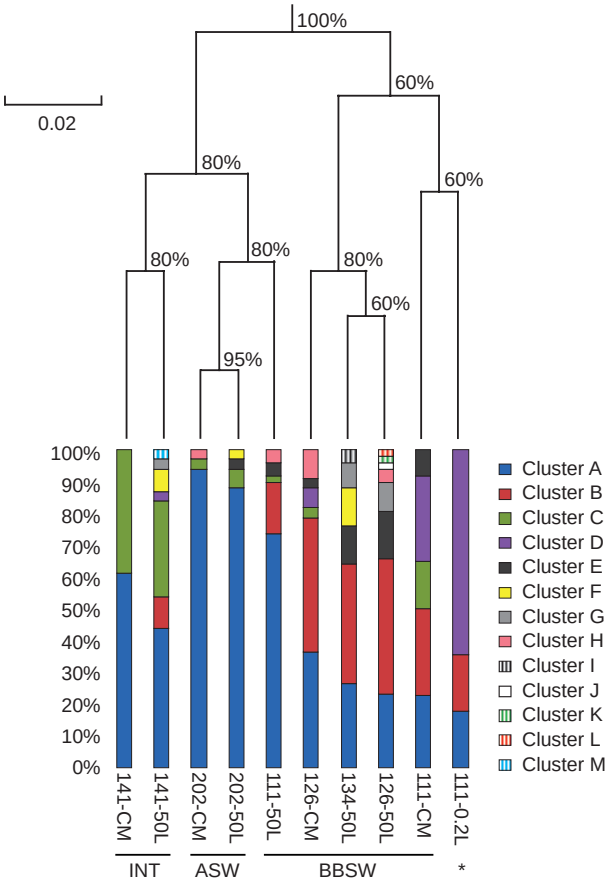

Fig. 4. Comparison of *Rhodobacterales* community structure in northern Baffin Bay. The top tree was constructed with the UniFrac jackknife clustering, based on the phylogeny of all *g5* clones in the 10 samples. The support value at each node is a percentage calculated from 1000 iterations, and the scale bar represents the UniFrac distance. The bar chart shows the proportional occurrence of each identified cluster in all samples, as indicated in the color legend. The sample identifiers and water mass are at the bottom of the figure (see Table 1 for details); *, not assigned to any water mass.

of the major active bacterial groups in northern Baffin Bay. This cluster is dominant in the ASW samples (Fig. 4), which are the coldest samples. The abundance and phylogeny of this cluster suggest that it is related to the RCA cluster that is often detected in large numbers from polar and subpolar oceans (Selje *et al.*, 2004; Buchan *et al.*, 2005; Wagner-Döbler & Biebl, 2006; Giebel *et al.*, 2009). The relative amounts of the RCA cluster have been shown to correlate with a number of different environmental variables such as prokaryote abundance, particulate matter, temperature, and phytoplankton pigments (Giebel *et al.*, 2009, 2011; Sperling *et al.*, 2012).

The sequences in OTU ii (cluster B) are most closely related to *Oceanibulbus indolifex*. Strains from this genus are described as mesophiles, with a documented lowest

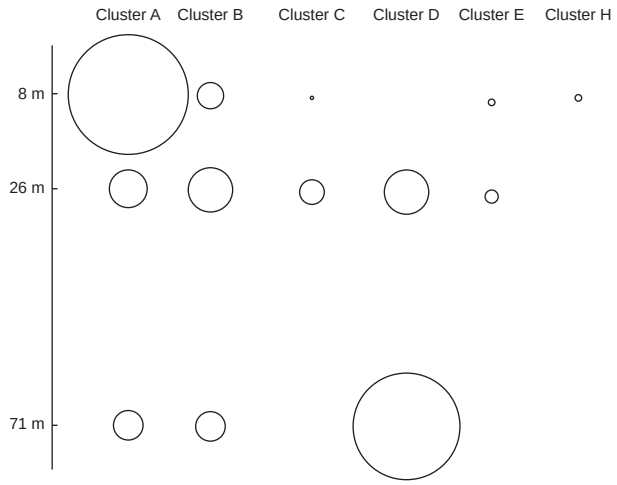

Fig. 5. Proportional occurrence of clusters in samples collected at three depths from station 111. The 50% surface light intensity depth was 8 m, the depth where chl *a* concentration was the highest was 26 m, and the 0.2% surface light intensity depth was 71 m. The diameter of each circle represents the relative proportional occurrence of each cluster in each sample.

growth temperature of 8 °C (Wagner-Döbler *et al.*, 2004). This OTU was absent from ASW clone library sequences and most of the clones (92%) representing this OTU were obtained from BBSW samples, which have the higher water temperatures compared to the other water masses. However, the proportional occurrence of this cluster was not significantly correlated with temperature (Table 3). OTU vii grouped with two *Sulfitobacter* strains at the species level ($\geq$ 95% identity). This is the highest similarity we found between clones and cultured representatives. However, OTU vii only contributed 2% of total clones and they were all found in surface water at relatively warm stations. This pattern could be explained by the temperature preferences of *Sulfitobacter* isolates, where the observed optimal growth temperature of cultured *Sulfitobacter* strains is ~20 °C (Sorokin, 1995), and perhaps, the low temperature of northern Baffin Bay limits their distribution. Understanding these distribution patterns is important in light of ongoing climate change, which will likely alter bacterial distributions in the oceans and thereby alter the represented microbial physiological activities.

The genome sequences available for isolates that affiliate with our dominant sequence groups (Fig. 2) show conservation of the GTA gene cluster (Newton *et al.*, 2010). In the limited cases where *Rhodobacterales* genomes do not contain contiguous GTA clusters, they appear to lack the entire set of genes and not just a portion of the cluster (Lang & Beatty, 2007; Biers *et al.*, 2008; Newton *et al.*, 2010; Lang *et al.*, 2012). Therefore, we predict that the *g5* amplicons represent strains with the genetic potential to produce functional GTA particles and

Table 4. Cultured organisms with *g5* sequences available that can be associated with *g5* library clone sequences.

Phylogenetic groups*	Reference cultured representatives
Cluster A	*Octadecabacter antarticus* 307
Cluster B, OTU ii	*Oceanibulbus indolifex* HEL-45
Cluster G	*Sulfitobacter* sp. EE-36, *Sulfitobacter* sp. NAS-14.1, *Roseobacter* sp. GAI101
OTU vii	*Sulfitobacter* sp. EE-36, *Sulfitobacter* sp. NAS-14.1

*Defined in Fig. 2.

our measure of *Rhodobacterales* diversity in this work also represents a measure of potential GTA diversity. However, lack of cultured representatives of the actual phylotypes identified in studies such as ours remains a problem in studying *Rhodobacterales* diversity and understanding the potential for GTA production in natural environments. For example, our cluster A is loosely related to the type strain of *Octadecabacter antarcticus* (Fig. 2; Table 4). The cluster A *g5* sequences share *c.* 89% sequence identity with this strain, but the cluster A clones share *c.* 98% identity with each other. Therefore, how well this type strain represents the most abundant *Rhodobacterales* in our samples and to what extent we can apply this strain's physiological characteristics to explain the observed biogeography are uncertain. Comparison of the genome contents of cultured strains and uncultivated marine Roseobacters has indicated there are important differences in their genomic repertoires (Luo *et al.*, 2012), further highlighting the need to better characterize these currently uncultured groups. As discussed above, cluster A likely corresponds to the RCA cluster, and a complete genome sequence for an RCA cluster representative, *Planktomarina temperata*, is reported to be in progress [(Giebel *et al.*, 2011); NCBI project ID 42833, accession PRJNA42833], which will help inform future studies. The overall lack of cultured representatives related to obtained *Rhodobacterales* sequences, as illustrated in Fig. 2, is a barrier to further explain the biogeography of this important group of bacteria.

Acknowledgements

We thank Alison Buchan for comments on the manuscript. YF and KFK were supported by fellowships from Memorial University School of Graduate Studies. This research was supported by funding from the Natural Sciences and Engineering Research Council (NSERC) to ASL and from NSERC and International Polar Year to RBR. We thank Steven Holland for provision of the rarefaction analysis tool through http://strata.uga.edu/software/index.html.

References

Allgaier M, Uphoff H, Felske A & Wagner-Döbler I (2003) Aerobic anoxygenic photosynthesis in *Roseobacter* clade bacteria from diverse marine habitats. *Appl Environ Microbiol* **69**: 5051–5059.

Baldwin AJ, Moss JA, Pakulski JD, Catala P, Joux F & Jeffrey WH (2005) Microbial diversity in a Pacific Ocean transect from the Arctic to Antarctic circles. *Aquat Microb Ecol* **41**: 91–102.

Baltar F, Arístegui J, Gasol J, Hernández-León S & Herndl G (2007) Strong coast-ocean and surface-depth gradients in prokaryotic assemblage structure and activity in a coastal transition zone region. *Aquat Microb Ecol* **50**: 63–74.

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

Bouvier TC & del Giorgio PA (2002) Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnol Oceanogr* **47**: 453–470.

Brinkhoff T, Bach G, Heidorn T, Liang L, Schlingloff A & Simon M (2004) Antibiotic production by a *Roseobacter* clade-affiliated species from the German Wadden Sea and its antagonistic effects on indigenous isolates. *Appl Environ Microbiol* **70**: 2560–2565.

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

Brinkmeyer R, Knittel K, Jurgens 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, Philip GK, Bunge JA, Smith MC, Bissett A, Lauro FM, Fuhrman JA & Donachie SP (2009) Microbial community structure in the North Pacific Ocean. *ISME J* **3**: 1374–1386.

Buchan A & González JM (2010) Roseobacter. *Handbook of Hydrocarbon and Lipid Microbiology* (Timmis KN, ed.), pp. 1335–1343. Springer, Berlin Heidelberg.

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

Buchan A, Hadden M & Suzuki MT (2009) Development and application of quantitative-PCR tools for subgroups of the *Roseobacter* clade. *Appl Environ Microbiol* **75**: 7542–7547.

Cunliffe M (2011) Correlating carbon monoxide oxidation with *cox* genes in the abundant Marine Roseobacter Clade. *ISME J* **5**: 685–691.

Dang H, Li T, Chen M & Huang G (2008) Cross-ocean distribution of *Rhodobacterales* bacteria as primary surface colonizers in temperate coastal marine waters. *Appl Environ Microbiol* **74**: 52–60.

Fu Y, MacLeod DM, Rivkin RB, Chen F, Buchan A & Lang AS (2010) High diversity of *Rhodobacterales* in the subarctic North Atlantic Ocean and gene transfer agent protein expression in isolated strains. *Aquat Microb Ecol* **59**: 283–293.

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL & Brown JH (2008) A latitudinal diversity gradient in planktonic marine bacteria. *P Natl Acad Sci USA* **105**: 7774–7778.

Galand PE, Potvin M, Casamayor EO & Lovejoy C (2010) Hydrography shapes bacterial biogeography of the deep Arctic Ocean. *ISME J* **4**: 564–576.

García-Martínez J & Rodríguez-Valera F (2000) Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I. *Mol Ecol* **9**: 935–948.

Ghiglione J-F, Galand PE, Pommier T *et al.* (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. *P Natl Acad Sci USA* **109**: 17633–17638.

Giebel HA, Brinkhoff T, Zwisler W, Selje N & Simon M (2009) Distribution of *Roseobacter* RCA and SAR11 lineages and distinct bacterial communities from the subtropics to the Southern Ocean. *Environ Microbiol* **11**: 2164–2178.

Giebel HA, Kalhoefer D, Lemke A, Thole S, Gahl-Janssen R, Simon M & Brinkhoff T (2011) Distribution of *Roseobacter* RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. *ISME J* **5**: 8–19.

Gilbert JA, Steele JA, Caporaso JG *et al.* (2012) Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298–308.

González JM & Moran MA (1997) Numerical dominance of a group of marine bacteria in the α-subclass of the class *Proteobacteria* in coastal seawater. *Appl Environ Microbiol* **63**: 4237–4242.

González JM, Simó R, Massana R, Covert JS, Casamayor EO, Pedrós-Alió C & Moran MA (2000) Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl Environ Microbiol* **66**: 4237–4246.

Good IJ (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* **40**: 237–264.

Gosink JJ, Herwig RP & Staley JT (1997) *Octadecabacter arcticus* gen. nov., sp. nov., and *O. antarcticus*, sp. nov., nonpigmented, psychrophilic gas vacuolate bacteria from polar sea ice and water. *Syst Appl Microbiol* **20**: 356–365.

Gram L, Grossart HP, Schlingloff A & Kiorboe T (2002) Possible quorum sensing in marine snow bacteria: production of acylated homoserine lactones by *Roseobacter* strains isolated from marine snow. *Appl Environ Microbiol* **68**: 4111–4116.

Green DH, Llewellyn LE, Negri AP, Blackburn SI & Bolch CJS (2004) Phylogenetic and functional diversity of the cultivable bacterial community associated with the paralytic shellfish poisoning dinoflagellate *Gymnodinium catenatum*. *FEMS Microbiol Ecol* **47**: 345–357.

Hale MS, Rivkin RB, Matthews P, Agawin NS & Li WKW (2006) Microbial response to mesoscale iron enrichment in the NE Subarctic Pacific: bacterial dynamics. *Deep Sea Res II* **53**: 2231–2247.

Hanson CA, Fuhrman JA, Horner-Devine MC & Martiny JB (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

Hobbie JE, Daley RJ & Jasper S (1977) Use of nuclepore filters for counting bacteria by fluorescence microscopy. *Appl Environ Microbiol* **33**: 1225–1228.

Kristensen DM, Mushegian AR, Dolja VV & Koonin EV (2010) New dimensions of the virus world discovered through metagenomics. *Trends Microbiol* **18**: 11–19.

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

Lang AS & Beatty JT (2010) Gene transfer agents and defective bacteriophages as sources of extracellular prokaryotic DNA. *Extracellular Nucleic Acids* (Kikuchi Y & Rykova EY, eds), pp. 15–24. Springer, Heidelberg.

Lang AS, Zhaxybayeva O & Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nat Rev Microbiol* **10**: 472–482.

Li WKW (1995) Composition of ultraphytoplankton in the central North-Atlantic. *Mar Ecol Prog Ser* **122**: 1–8.

Li WKW & Dickie PM (2001) Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry* **44**: 236–246.

Lozupone C & Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

Luo H, Löytynoja A & Moran MA (2012) Genome content of uncultivated marine *Roseobacters* in the surface ocean. *Environ Microbiol* **14**: 41–51.

Marie D, Brussaard CPD, Thyrhaug R, Bratbak G & Vaulot D (1999) Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl Environ Microbiol* **65**: 45–52.

Martens T, Gram L, Grossart HP, Kessler D, Muller R, Simon M, Wenzel SC & Brinkhoff T (2007) Bacteria of the *Roseobacter* clade show potential for secondary metabolite production. *Microb Ecol* **54**: 31–42.

Martiny JB, Bohannan BJ, Brown JH *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB & Paul JH (2010) High frequency of horizontal gene transfer in the oceans. *Science* **330**: 50.

McDaniel LD, Young EC, Ritchie KB & Paul JH (2012) Environmental factors influencing Gene Transfer Agent (GTA) mediated transduction in the subtropical ocean. *PLoS ONE* **7**: e43506.

Melling H, Gratton Y & Ingram G (2001) Ocean circulation within the North Water polynya of Baffin Bay. *Atmosphere–Ocean* **39**: 301–325.

Moran MA, Belas R, Schell MA *et al.* (2007) Ecological genomics of marine *Roseobacters*. *Appl Environ Microbiol* **73**: 4559–4569.

Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA & Giovannoni SJ (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–810.

Morris RM, Frazar CD & Carlson CA (2012) Basin-scale patterns in the abundance of SAR11 subclades, marine *Actinobacteria* (OM1), members of the *Roseobacter* clade and OCS116 in the South Atlantic. *Environ Microbiol* **14**: 1133–1144.

Motard-Côté J, Levasseur M, Scarratt MG *et al.* (2012) Distribution and metabolism of dimethylsulfoniopropionate (DMSP) and phylogenetic affiliation of DMSP-assimilating bacteria in northern Baffin Bay/Lancaster Sound. *J Geophys Res – Oceans* **117**: C00G11.

Nemergut DR, Costello EK, Hamady M *et al.* (2011) Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13**: 135–144.

Newton RJ, Griffin LE, Bowles KM *et al.* (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME J* **4**: 784–798.

Parsons TR, Maita Y & Lalli CM (1984) *A Manual of Chemical and Biological Methods for Seawater Analysis*. Pergamon, Oxford.

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

Pinhassi J, Winding A, Binnerup SJ, Zweifel UL, Riemann B & Hagström Å (2003) Spatial variability in bacterioplankton community composition at the Skagerrak-Kattegat front. *Mar Ecol Prog Ser* **255**: 1–13.

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

Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P, Tunlid A & Hagstrom A (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16**: 867–880.

Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA & Pernthaler J (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11**: 2078–2093.

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

Seyedsayamdost MR, Case RJ, Kolter R & Clardy J (2011) The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nat Chem* **3**: 331–335.

Shannon CE & Weaver W (1949) *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL.

Shiba T (1991) *Roseobacter litoralis* gen. nov., sp. nov., and *Roseobacter denitrificans* sp. nov., aerobic pink-pigmented bacteria which contain Bacteriochlorophyll *a*. *Syst Appl Microbiol* **14**: 140–145.

Simon M & Azam F (1989) Protein-content and protein-synthesis rates of planktonic marine-bacteria. *Mar Ecol Prog Ser* **51**: 201–213.

Sorokin DY (1995) *Sulfitobacter pontiacus* gen. nov., sp. nov. – a new heterotrophic bacterium from the Black Sea, specialized on sulfite oxidation. *Microbiology* **64**: 295–305.

Sperling M, Giebel H, Rink B, Grayek S, Staneva J, Stanev E & Simon M (2012) Differential effects of hydrographic and biogeochemical properties on the SAR11 clade and *Roseobacter* RCA cluster in the North Sea. *Aquat Microb Ecol* **67**: 25–34.

Staley JT & Gosink JJ (1999) Poles apart: biodiversity and biogeography of sea ice bacteria. *Annu Rev Microbiol* **53**: 189–215.

Stanton TB (2007) Prophage-like gene transfer agents-Novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe* **13**: 43–49.

Tamura K, Nei M & Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *P Natl Acad Sci USA* **101**: 11030–11035.

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.

Taniguchi A & Hamasaki K (2008) Community structures of actively growing bacteria shift along a north-south transect in the western North Pacific. *Environ Microbiol* **10**: 1007–1017.

Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Vila-Costa M, Rinta-Kanto JM, Sun S, Sharma S, Poretsky R & Moran MA (2010) Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. *ISME J* **4**: 1410–1420.

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

Wagner-Dobler I, Rheims H, Felske A, El-Ghezal A, Flade-Schroder D, Laatsch H, Lang S, Pukall R & Tindall BJ (2004) *Oceanibulbus indolifex* gen. nov., sp. nov., a North Sea alphaproteobacterium that produces bioactive metabolites. *Int J Syst Evol Microbiol* **54**: 1177–1184.

Webster NS, Negri AP, Munro MMHG & Battershill CN (2004) Diverse microbial communities inhabit Antarctic sponges. *Environ Microbiol* **6**: 288–300.

Yao D, Buchan A & Suzuki MT (2011) *In situ* activity of NAC11-7 roseobacters in coastal waters off the Chesapeake Bay based on *ftsZ* expression. *Environ Microbiol* **13**: 1032–1041.

Zhao Y, Wang K, Budinoff C, Buchan A, Lang A, Jiao N & Chen F (2009) Gene transfer agent (GTA) genes reveal diverse and dynamic *Roseobacter* and *Rhodobacter* populations in the Chesapeake Bay. *ISME J* **3**: 364–373.

Supporting Information

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

Fig. S1. Temperature and salinity for the stations visited during the 2008 SOLAS project.
Fig. S2. Rarefaction curves for the 10 *g5* clone libraries.
Table S1. Sequence identification information for the *g5* clones sequenced in this study, GenBank accession numbers KC111451–KC111767.