High diversity of *Rhodobacterales* in the subarctic North Atlantic Ocean and gene transfer agent protein expression in isolated strains

Yunyun Fu¹,*, Dawne M. MacLeod¹,*, Richard B. Rivkin², Feng Chen³, Alison Buchan⁴, Andrew S. Lang¹,**

¹Department of Biology, Memorial University of Newfoundland, 232 Elizabeth Ave., St. John’s, Newfoundland A1B 3X9, Canada
²Ocean Sciences Centre, Memorial University of Newfoundland, Marine Lab Road, St. John’s, Newfoundland A1C 5S7, Canada
³Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 236-701 East Pratt St., Baltimore, Maryland 21202, USA
⁴Department of Microbiology, University of Tennessee, M409 Walters Life Sciences, Knoxville, Tennessee 37914, USA

ABSTRACT: Genes encoding gene transfer agent (GTA) particles are well conserved in bacteria of the order *Rhodobacterales*. Members of this order are abundant in diverse marine environments, frequently accounting for as much as 25% of the total bacterial community. Conservation of the genes encoding GTAs allows their use as diagnostic markers of *Rhodobacterales* in biogeographical studies. The first survey of the diversity of *Rhodobacterales* based on the GTA major capsid gene was conducted in a warm temperate estuarine ecosystem, the Chesapeake Bay, but the biogeography of *Rhodobacterales* has not been explored extensively. This study investigates *Rhodobacterales* diversity in the cold subarctic water near Newfoundland, Canada. Our results suggest that the subarctic region of the North Atlantic contains diverse *Rhodobacterales* communities in both winter and summer, and that the diversity of the *Rhodobacterales* community in the summer Newfoundland coastal water is higher than that found in the Chesapeake Bay, in either the summer or winter. Approximately one-third of GTA sequences retrieved from the subarctic waters were most closely related to those from bacteria isolated from sea ice or cold regions. Distinguishable diversity patterns were found between the temperate and subarctic waters, providing further support for niche adaptation of specific *Rhodobacterales* members to unique environments. We also demonstrate that a number of *Rhodobacterales* strains, from both the subarctic and temperate locations, express the GTA major capsid protein. This provides robust evidence that the widespread conservation of GTA genes in the *Rhodobacterales* may result in the production of functionally similar and active GTA systems in these bacteria in different environments.

KEY WORDS: *Rhodobacterales* · Gene transfer agent · GTA · Major capsid protein

INTRODUCTION

Estimates in marine microbial populations suggest members of the α-proteobacterial order *Rhodobacterales*, specifically the *Roseobacter* clade, can comprise upwards of 25% of total marine bacterioplankton (Buchan et al. 2005). All complete *Rhodobacterales* genome sequences contain gene transfer agent (GTA) gene clusters (Lang & Beatty 2007, Biers et al. 2008). GTAs are bacteriophage-like particles that package and transfer genomic DNA between bacterial cells in a process analogous to transduction.
GTAs are distinguished from transducing phages in several ways, including the apparently random and obligate packaging of fragments of the producing cell’s genome, and the fact that the size of DNA packaged in the particles is insufficient to encode the full GTA structure (Lang & Beatty 2007, Stanton 2007). This mechanism of genetic exchange has been identified in diverse prokaryotic species: the purple non-sulfur α-proteobacterium *Rhodobacter capsulatus* (Marrs 1974), the sulfate-reducing δ-proteobacterium *Desulfovibrio desulfuricans* (Rapp & Wall 1987), the spirochete *Brachyspira hyodysenteriae* (Humphrey et al. 1997), the methanogenic archaeon *Methanococcus voltae* (Bertani 1999), and the marine α-proteobacterium *Ruegeria pomeroyi* (Biers et al. 2008). With the exception of the *R. pomeroyi* GTA (Biers et al. 2008), all of these particles structurally resemble tailed phages (reviewed in Stanton 2007).

The role of GTAs in natural environments is not yet understood, but they exist in the above-mentioned (and presumably other) phylogenetically diverse prokaryotes, suggesting this mode of DNA transfer may be important in shaping some microbial genomes and communities. An outcome of GTA production for the producing species is homologous recombination. Therefore GTA activity could provide benefits to the producing species as a whole (Vos 2009), even if it is detrimental to the individual producing cell that presumably must lyse to allow GTA release. The release of GTAs is known to occur by lysis only in *Brachyspira hyodysenteriae* (Humphrey et al. 1997, Matson et al. 2005), but it is anticipated that cell lysis is the likely mechanism of release for all GTAs because there is no known release of tailed phage particles from cells other than through lysis.

The *Rhodobacter capsulatus* GTA (RcGTA) gene cluster comprises 16 open reading frames (ORFs) in ~14 kb on the bacterial chromosome (Lang & Beatty 2000, Lang & Beatty 2007). Several of these ORFs have recognizable homology to known bacteriophage genes, such that the cluster appears to be an isolated head- and tail-encoding unit. Of these 16 ORFs, 3 are particularly well conserved between different species (Lang et al. 2002): *g2*, predicted to encode the large terminase protein, *g5*, encoding the major capsid protein, and *g9*, predicted to encode the major tail protein (Lang & Beatty 2000, Lang & Beatty 2007). Due to the conservation of GTA genes in the *Rhodobacter* order, the GTA major capsid protein-encoding gene, *g5*, has been used as a marker to estimate *Rhodobacter* diversity in estuarine microbial communities (Zhao et al. 2009). This work demonstrated that there was a distinct temporal and spatial pattern of *Rhodobacter* diversity in the Chesapeake Bay and that *g5* sequence diversity was higher in these natural communities than appreciated from previous isolates.

We have exploited the GTA gene conservation to explore the diversity of *Rhodobacterales* in microbial communities in the subarctic western North Atlantic Ocean. The results presented here demonstrate that the cold productive subarctic waters of coastal Newfoundland (Canada) harbor a remarkably diverse collection of *Rhodobacterales*, and approximately one-third of the GTA gene sequences retrieved from these waters were most similar to those from bacteria isolated from sea ice or cold regions. The propensity of the characterized representatives of ecologically important *Rhodobacterales* bacteria to display the genetic potential for production of GTAs warrants further investigation of this property. Therefore, we have also cultured several strains from this coastal environment and demonstrated GTA capsid protein expression amongst these, as well as in strains recovered from the Chesapeake Bay. These findings show that, not only the genetic potential for GTA production, but also the expression of the GTA capsid protein is widespread, both taxonomically and biogeographically, among members of the *Rhodobacterales*. This research provides further evidence that GTA-mediated gene transfer could be an important phenomenon in natural microbial communities.

**MATERIALS AND METHODS**

**Sample collection.** Microbial communities were collected onto 0.2 µm polycarbonate filters (Millipore) from 500 ml surface water samples collected at Tappar’s Cove (47°39’54”N, 52°43’38”W) on July 24, 2007, and at Logy Bay (47°38’14”N, 52°39’36”W) on December 11, 2008. These are adjacent locations that experience very similar climate and hydrographic forcing, and the inshore branch of the Labrador Current influences both bays. DNA was extracted from the filters as previously described (Kan et al. 2006). Filters were thawed and each one was combined with 2 ml pre-lysis buffer (0.1 M Tris-HCL; 0.1 M EDTA; 0.8 M sucrose; pH 8) and 10 µl lysozyme (200 µg ml⁻¹) and incubated at 37°C for 30 min. This incubation 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 then sequentially extracted with phenol-chloroform-isooamyl alcohol (25:24:1) and chloroform-isoamyl alcohol (24:1), followed by isopropanol precipitation. Following a wash with 70% ethanol, the pellets were air-dried and dissolved in 200 µl Tris-EDTA (TE) buffer.
Water samples for enumerating bacterial abundance were fixed with 0.2 µm filtered formaldehyde (final concentration 3.7%) and filtered onto 0.2 µm polycarbonate membrane filters. Cells were stained with 1 µg ml\(^{-1}\) DAPI (4',6-diamidino-2-phenylindole) and counted by epifluorescence microscopy at 1250x magnification using UV light excitation. For each sample, at least 1000 cells were enumerated.

**PCR amplification, cloning and sequencing of GTA g5 genes.** Degenerate primers for amplification of GTA g5 genes (encoding the major capsid protein) described previously (Zhao et al. 2009) were used in the present study: MCP-109F, 5'-GGC TAY CTT GTS GAT CCS CAR AC-3' and MCP-368R, 5'-TAG AAC AGS ACR TGS GGY TTK GC-3'. Target DNA was amplified in a single round of PCR in 25 µl volumes containing 0.5 µM of each primer, 3% DMSO (v/v), 200 µM dNTPs, 1× Phusion Buffer (NEB), and 0.02 U µl\(^{-1}\) of Phusion Hot Start polymerase (NEB). Thermal cycling conditions were as follows: 30 s at 98°C, followed by 35 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 30 s, and a final extension step at 72°C for 7 min. PCR products were purified using a Qiagurt PCR Clean-up kit (Qiagen) and 3' adenine overhangs were added by incubation with 1 U of Taq polymerase (NEB) at 72°C for 10 min in 1× Taq buffer (NEB). PCR products were cloned into the pGEM-T Easy Vector System (Promega) according to the manufacturer's recommendations, and clones containing appropriately sized inserts were sequenced at either the Genomics and Proteomics Facility at Memorial University (St. John's, Canada) or the Centre for Applied Genomics (Toronto, Canada) using the M13 primers that bind to the cloning vector.

**Isolation and characterization of Rhodobacterales.** A sub-sample of the seawater collected at Logy Bay on December 11, 2008 for community analysis (see 'Sample collection') was used for isolation of Rhodobacterales bacteria. Seawater was spread directly onto sterile Petri plates that contained dimethylsulfonyl propionate (DMSP) as the sole carbon source. This enrichment is designed to select for bacteria of the Rhodobacterales order, which often display the ability to use DMSP as a sole carbon and energy source. Several colonies on these second DMSP plates were then replated and maintained on a complex medium (YTSS; 4 g l\(^{-1}\) tryptone, 2.5 g l\(^{-1}\) yeast extract, 15 g l\(^{-1}\) sea salts; Sigma-Aldrich). Each culture was then grown in liquid broth at 25°C for 1 d and DNA was extracted using the Puregene DNA Purification kit (Qiagen). The DNA from these isolates was used as templates for PCR with the g5 primers (Zhao et al. 2009) as described in the previous section, purified and directly sequenced using the g5 primers. The primers 27F (5'-AGA GTT TGA TCM TGG CTC AG-3') (Giovannoni 1991) and Roseo536R (5'-CAA CGC TAA CCC CCT CCG-3') (Brinkmeyer et al. 2000) were used to amplify a ~500 bp region of the 16S rRNA gene. The 16S rDNA PCR reaction components were the same as for the g5 reactions with the following thermocycling conditions: 30 s at 98°C, followed by 35 cycles of 98°C for 10 s, 58°C for 30 s, and 72°C for 30 s, and a final extension step at 72°C for 7 min. Resulting PCR products were purified as described above and sequenced using these same 16S primers. If this sequence identified the isolate as a member of the Rhodobacterales, the nearly complete 16S rRNA gene was amplified and sequenced using the primers 27F and 1522R (5'-AAG GAG GTG ATC CAN CCR CA-3') (Giovannoni 1991). The PCR reaction components and thermocycling conditions were as described above with the extension step increased to 60 s.

**Sequence analyses.** Sequences were edited to remove vector and primer sequences and aligned with Geneious v.3.8.5 (Biomatters). Sequences were assigned to operational taxonomic units (OTUs), within each sample library, using 97% identity as the criterion, as conducted previously (Zhao et al. 2009). The proportion of sequence diversity that had been discovered within each clone library was estimated by calculation of the percent coverage (C) using the formula \(C = 1 - (N/n) \times 100\), where N is the number of unique sequences (i.e. those found only once in that sample), and n is the total number of clones sequenced (Raven-schlag et al. 1999). The Shannon diversity index (\(H'\)) was calculated for the samples based on the OTU detection frequencies using the PAST software package (Hammer et al. 2001). Rarefaction analysis of each sample was conducted using Analytical Rarefaction v1.4 (www.uga.edu/strata/software/Software.html) to assess what proportion of the sample diversity (at least the portion that was successfully amplified) had been accounted for.

For phylogenetic analyses, alignments of 16S rDNA sequences were performed using ClustalW within the program MEGA4 (Tamura et al. 2007), evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004), and the evolutionary pattern was inferred by the Neighbor-Joining method (Saitou & Nei 1987) using the pair-wise deletion of gaps option. A similar approach was used for analysis of the g5 sequences; however, these
nucleotide sequences were aligned according to the corresponding aligned amino acid sequences. Phylogenetic analyses were conducted for the environmental sequences reported here, sequences from the Chesapeake Bay (Zhao et al. 2009), and sequences from complete *Rhodobacterales* genomes that were available in GenBank. Accession numbers for previously published sequences are provided in Table S1, Supplement 1 (at www.int-res.com/articles/suppl/a059p283_app.pdf). The *g5* clone OTU sequences described in the present study have been submitted to GenBank and assigned accession numbers FJ872412 to FJ872476. Sequences from the isolates described in this work have been submitted to GenBank and assigned accession numbers GQ433706 to GQ433715.

Predicted molecular weights of capsid proteins were calculated based on the amino acid sequences using Geneious v.3.8.5 (Biomatters).

**Western blotting.** The control *Rhodobacter capsulatus* cultures were grown photosynthetically in YPS medium (Wall et al. 1975). *R. capsulatus* SB1003 (Yen & Marrs 1976) is a GTA-producing strain and *R. capsulatus* A1 does not produce the GTA capsid protein due to a Tn5 insertion in the *g5* gene (J. T. Beatty pers. comm.). The *Roseobacter* strains were grown in liquid YTSS broth at 25°C for 3 d. For cell samples, cells were pelleted by centrifugation at 17 000 g for 2 min, the medium was removed and the cells were resuspended in an equal volume of TE buffer. For supernatant samples, cells were removed by 2 rounds of centrifugation at 17 000 g for 2 min, with a sub-sample of the supernatant removed to a new tube after each centrifugation step. Two volumes of the cells or final culture supernatant were mixed with 1 volume 3× sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer (NEB) and heated to 98°C for 5 min. The samples were subjected to SDS-PAGE on 10% polyacrylamide gels and transferred to a nitrocellulose membrane. The membrane was blocked with 5% (w/v) skim milk in Tris-buffered saline Tween-20 (TBST) (20 mM Tris and 137 mM NaCl, pH 7.5; 0.1% [v/v] Tween 20) and then incubated with primary antibody against the *Rhodobacterales* GTA major capsid protein (AS08 365; Agrisera AB) in TBST at 4°C overnight. After washing, the membrane was incubated with peroxidase-conjugated goat anti-rabbit IgG (Santa Cruz Biotechnology), and specific bands were detected by enhanced chemiluminescence with the SuperSignal West Femto Reagent (Pierce Biotechnology).

The antibody targeting the *Rhodobacterales* GTA major capsid protein is a polyclonal antibody that was raised against a synthetic peptide corresponding to a region of the GTA major capsid protein that is conserved at >90% identity in all currently available *Rhodobacterales* sequences (Agrisera AB). An antibody neutralization experiment was used to confirm that we were detecting bands specifically recognized by the anti-GTA primary antibodies. For the antibody neutralization experiment, the primary antibody was incubated with an approximately 100-fold molar excess of the immunizing peptide in TBST for 2 h at room temperature. The control antibody (non-neutralized) was similarly pre-incubated in TBST for 2 h, but without added peptide. The 2 antibody solutions were added to duplicate membranes with assorted *Rhodobacterales* strains/species.

**RESULTS AND DISCUSSION**

High diversity of *Rhodobacterales* in the western North Atlantic Ocean

We analyzed GTA *g5* clone libraries generated from 2 coastal western North Atlantic Ocean surface water samples, one collected in summer and one collected in winter. The sequencing of 172 clones from the 2 libraries, 127 from the summer and 45 from the winter, led to the discovery of a diverse collection of *Rhodobacterales* communities in the sub-arctic water near Newfoundland. Fifty-nine OTUs were defined as sequences sharing ≥97% nucleotide identity. Of these 59 OTUs, 49 were found only in the summer sample, 5 were found only in the winter sample and 5 were found in both. There is a wide range of homology between the OTUs, with percent nucleotide identities ranging from 44 to 97%, but the vast majority are between 50 and 70% identical (Supplement 2 at www.int-res.com/articles/suppl/a059p283_app.xls). Pair-wise comparisons of all *Rhodobacterales* (partial) *g5* sequences show that the most dissimilar are 53 and 42% identical at the protein and nucleotide levels, respectively. It is possible that the value of 97% identity is not a meaningful measure for these gene sequences, as we know little about the relationship between sequence variation and function for this gene. However, when the cutoff is reduced to 90%, significant gene diversity is still evident (48 total OTUs across both samples).

The summer microbial community had a more diverse population of *Rhodobacterales* compared with the winter (Table 1, Fig. S1 in Supplement 1 at www.int-res.com/articles/suppl/a059p283_app.pdf). Furthermore, the Shannon index in the summer Newfoundland coastal water (*H'*= 3.71) was higher than the highest found in the Chesapeake samples (*H'*= 2.39 for the winter upper bay sample) (Zhao et al. 2009). Heterotrophic bacteria in the subarctic area of North Atlantic are known to grow very slowly, even during phytoplankton blooms (Pomeroy & Deibel 1986), with growth rates as low as 0.002 divisions per day.
have discovered clearly group with previously these genes illustrates that some of the sequences we possible reasons for these observations.ampling program will be required to validate and address 2005, Rink et al. 2007). A more comprehensive sam-

Another factor that is expected to influence the cycle between ~1 and 26°C cannot be discounted. down the estuary, although the annual temperature fluctuation (–1 to 14°C) is similar to the New-

Temperature fluctuation (–1 to 14°C) is thought to be a major environmental factor (reviewed in Rivkin et al. 1996); low temperature (i.e. <4°C) is thought to be a major environmental factor causing low production by heterotrophic bacteria. It is intriguing to learn that the overall diversity of the Rhodobacterales community, as inferred from g5 gene sequences, in the subarctic coastal water appears higher than that in the temperate estuary. It is known that some Roseobacter bacteria are adapted to cold environments (Brinkhoff et al. 2008). In one study, based on clone library representation, it was estimated that members of the Roseobacter group could make up >50% of total bacterial communities in sub-Antarctic waters (Prabagaran et al. 2007), where the annual temperature fluctuation (~1 to 14°C) is similar to the Newfoundland coastal water. In Newfoundland, winter water temperatures are low (~1.5 to 1°C) while summer temperatures typically reach 14 to 16°C in late August and September. This contrasts with Chesapeake Bay, where river flow from the Susquehanna and the other major tributaries (Choptank, Chester, Potomac, James and Rappahanock Rivers) is one of the main environmental forcing factors, leading to a salinity gradient down the estuary, although the annual temperature gradient between ~1 and 26°C cannot be discounted. Another factor that is expected to influence the Rhodobacterales communities is the spring phytoplankton bloom, which typically occurs between late February and late March in the Chesapeake Bay and between mid-April and late May on the east coast of Newfoundland, because the abundance of these bacteria is positively influenced by phytoplankton (Buchan et al. 2005, Rink et al. 2007). A more comprehensive sampling program will be required to validate and address possible reasons for these observations.

A phylogenetic comparison of the sequences from these genes illustrates that some of the sequences we have discovered clearly group with previously described species or strains, while others are more highly divergent, and do not group closely with previously reported GTA capsid sequences (Fig. 1). We found a large number of sequences (10 OTUs, 19 clones), from both summer and winter samples, which cluster with the capsid sequence from Octadecabacter antarcticus, an isolate recovered from sea ice (Gosink et al. 1997). These 10 OTUs were distributed into 4 independent branches within the Octadecabacter clade, suggesting the presence of microdiversity within this group. Similarly, there were 10 OTUs (from 51 clones) that cluster most closely with the sequence from the North Sea bacterium Oceanibulbus indolifex (Wagner-Dohler et al. 2004). These findings provide evidence that organisms closely related to these 2 Roseobacter species may be prevalent in the sub-arctic North Atlantic; sequences that group closely with these species were not recovered from the Chesapeake Bay samples. Another group that appears abundant in our samples, and in the Chesapeake Bay, is related to the genus Loktanella (potentially 12 OTUs from 34 clones); however, the deep branches connecting these sequences are not well supported (Figs. 1 & S2 in Supplement 1 at www.int-res.com/articles/suppl/a059p283_app.pdf). We found 2 clone sequences that fall within the Sulfitobacter clade, and our culturing approach recovered isolates in this group (see next section). Where possible, we have grouped multiple clone OTUs into clusters (Fig. 1, labeled 1 to 7) that do not show well-supported relationships to any of the known strains/species; there are 3 remaining OTU sequences that are scattered amongst the various Roseobacter groups, again without strongly supported relationships to any of the known strains/species. Another observation is that relatively few sequences were identified in the present study that are closely related to Roseovarius, Phaeobacter, Oceanicola, or Ruegeria strains/species (Fig. 1), most of which have been isolated from more southerly environments.

The previously established congruency between 16S rDNA and g5-based phylogenies (Lang & Beatty 2007, Biers et al. 2008) does not necessarily hold true when a subset of the gene information (<750 positions) is used, as demonstrated here. For example, the Rhodobacter capsulatus g5 sequence falls within the Roseobacter group as opposed to grouping with the other Rhodobacter sequences (i.e. R. sphaeroides and Paracoccus denitrificans; Fig. 1). Thus, analysis of the amplified g5 product does not always accurately reflect the expected relationships. This same reasoning presumably explains why some groupings on the g5 phylogenetic tree are not well supported in the bootstrap analysis, especially the deeper branches. However, the g5 marker is still a valuable target for evaluating Rhodobacterales diversity and community

<table>
<thead>
<tr>
<th>Parametersa</th>
<th>July 2007</th>
<th>December 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial abundance (cells ml−1)b</td>
<td>5.4 × 10^5</td>
<td>5.5 × 10^5</td>
</tr>
<tr>
<td>Number of clones sequenced</td>
<td>127</td>
<td>45</td>
</tr>
<tr>
<td>Number of OTUs</td>
<td>54</td>
<td>10</td>
</tr>
<tr>
<td>% coverage (C)</td>
<td>80</td>
<td>93</td>
</tr>
<tr>
<td>Shannon-Wiener index (H')</td>
<td>3.71</td>
<td>1.59</td>
</tr>
</tbody>
</table>

aDetails of calculations are given in the ‘Materials and methods’ section
bCoefficient of variation 18.2% for July and 13.5% for December

Table 1. Characteristics of the samples used in this study and the resulting clone libraries. OTUs: operational taxonomic units.
Fig. 1. Phylogenetic tree of gene transfer agent (GTA) capsid gene sequences. Sequences amplified from the western North Atlantic microbial communities in this work are represented by yellow boxes for sequences from July 2007 and blue boxes for sequences from December 2008. Isolates identified in this study (DMLB-w4.2, DMLB-w6, DMLB-w8, DMLB-w12.1, and DMLB-w13) are highlighted in pink. Dark green boxes represent sequences from clones and light green boxes represent strains from Chesapeake Bay (Zhao et al. 2009). For simplicity of presentation, some clades that contain only clone sequences are collapsed, and the origins of the sequences in the collapsed clades are indicated using the colors as above. The clusters labelled 1–7 do not show well-supported relationships to any of the known strains/species. Numbers in the boxes represent the number of collapsed operons. The neighbor-joining tree is based on the aligned nucleotide sequences corresponding to the g5 amplicon (747 positions including gaps). Bootstrap values (percentages based on 10,000 replicates) are shown for the deepest branch points supported >60%. Scale bar indicates the number of substitutions per site. The g5 sequence from Brucella abortus was used as the outgroup for the tree. Accession numbers for previously published sequences and pairwise identity values are available in (Table S1, Supplement 1 and in, Supplement 2 (at www.int-res.com/articles/suppl/a059p283_app.pdf). The neighbor-joining tree is based on the aligned nucleotide sequences corresponding to the g5 amplicon (747 positions including gaps). Bootstrap values (percentages based on 10,000 replicates) are shown for the deepest branch points supported >60%. Scale bar indicates the number of substitutions per site. The g5 sequence from Brucella abortus was used as the outgroup for the tree. Accession numbers for previously published sequences and pairwise identity values are available in (Table S1, Supplement 1 and in, Supplement 2 (at www.int-res.com/articles/suppl/a059p283_app.pdf)

composition because of the ability to amplify these sequences from diverse isolates and uncultured representa-
tives within the order. Furthermore, the currently available evidence indicates that all members of this order contain this gene (Lang & Beatty 2007, Biers et al. 2008).

Despite the apparent universal GTA cluster conservation in Rhodobacterales genomes and the apparent high abundance of these bacteria in marine communities, GTA genes are generally not prevalent in marine metagenomes, particularly those derived from the Global Ocean Survey (GOS), the most comprehensive survey of marine surface waters to date. As previously suggested, this may be due to metagenome sampling biases or the characterized Rhodobacterales isolates may not be representative of strains found in natural communities (Biers et al. 2008). The latter explanation is not well supported; we have demonstrated here and in previous work (Zhao et al. 2009) an ability to consistently recover novel g5 gene sequences from natural systems. Thus, we support the former explanation, as the overwhelming majority of the GOS samples were processed through 0.8 µm filters (Venter et al. 2004, Rusch et al. 2007) that would remove larger, particle-associated or aggregated bacteria and thereby remove many Rhodobacterales. Indeed, a recent analysis of the GOS metagenomic data illustrates this bias (Biers et al. 2009).

Isolation of Roseobacter strains from the western North Atlantic and evaluation of GTA protein expression

We have cultured 5 independent isolates from the western North Atlantic samples that were identified as belonging to the Rhodobacterales, and specifically within the Roseobacter group (Fig. 2). Four of these (DMLB-w4.2, DMLB-w6, DMLB-w12.1 and DMLB-w13) are very closely related to one another and known strains in the Sulfitobacter genus, based on the g5 and 16S rRNA gene sequence information (Figs. 1 & 2). We also recovered 2 OTUs (representing 5 clones) that group closely with these Sulfitobacter sequences. Sulfitobacter isolates were also common in mid-Atlantic water during the winter (Zhao et al. 2009). The other isolate (DMLB-w8) falls within the Loktanella clade, based on analysis of 16S rDNA sequences (Fig. 2), although this relationship is not strongly supported by the partial g5 sequences (Fig. 1). Our culturing regime used an incubation temperature of 20°C, which would favor the growth of mesophilic strains and select against psychrophiles. This may explain why we did not recover isolates from the Oceanibulbus or Octadecabacter clades.

We used an antibody that is specific for the GTA capsid protein from Rhodobacterales to examine GTA protein expression in a panel of 12 phylogenetically diverse Rhodobacterales isolates from several different environments (Fig. 3), including the known GTA-producing species Rhodobacter capsulatus and Ruegeria pomeroyi. This is the first application of this tool for examining GTA protein expression. These antibodies allow rapid and accurate assessment of potential GTA production in cultured representatives, and will be invaluable for future studies aimed at studying the regulation of GTA production. The western blots showed that 6 of the other 10 strains we examined showed expression of the GTA capsid protein, and we were able to detect the protein in the supernatants of cultures from 4 of these. Four of the 5 Newfoundland isolates (from the Sulfitobacter clade) showed expression of a GTA protein, and this was detected in the supernatants for 2 of these, DMLB-w12.1 and DMLB-w13 (Fig. 3A). The 3 most closely related Newfoundland strains (DMLB-w4.2, DMLB-w6 and DMLB-w13; Fig. 1) showed 2 positive bands inside the cells, but only the lower of these is the result of a specific reaction with the primary antibody. Two isolates recovered from the Chesapeake Bay (CB1005 and CB1023) also gave positive results for the GTA capsid protein (Fig. 3B). The protein was not detected in Phaeobacter sp. Y4I, Citreicella sp. SE45, or isolates DMLB-w8 and
CB1040, both of which potentially group in the *Loktanella* clade. The predicted capsid proteins from all these strains contain the immunogenic sequence, and so the lack of detection in these strains could be due to several other reasons. The expression of the protein in the growth conditions used may be below the detection limit for this method in these strains; it is known that the amount of GTA produced by *R. capsulatus* (Solioz et al. 1975) and *R. pomeroyi* (Biers et al. 2008) is dependent on growth conditions. It is also possible that these strains may never express the GTA genes. However, it appears that GTA production is a property of bacteria in diverse genera of the *Rhodobacterales* order (Fig. 2). Work is underway to genetically demonstrate GTA-mediated genetic exchange by these strains.

We conducted peptide-blocking experiments to ensure that we were detecting bands specifically recognized by the antibodies raised against the conserved GTA peptide sequence. We used cell samples from a collection of the *Rhodobacterales* strains/species and incubated 2 equivalent membranes separately with either the anti-GTA antibody or the anti-GTA antibody baited 2 equivalent membranes separately with either the anti-GTA antibody or the anti-GTA antibody that had been pre-incubated with the immunizing peptide. This showed that the bands observed in *Rhodobacter capsulatus*, DMLB-w12.1, *Ruegeria pomeroyi* DSS-3, and CB1005 are specific (Fig. 3C). One of the 2 bands in DMLB-w6 is specific, while the other is not; close inspection of these lanes shows it is the lower MW band that is specific (Fig. 3S, in Supplement 1 at www.int-res.com/articles/suppl/a059p283_app.pdf).

The *Rhodobacter capsulatus* major capsid protein is proteolytically cleaved on the C-terminal side of a lysine residue to remove the first 103 amino acids and generate the mature protein found in the RcGTA structure that begins with alanine (Lang & Beatty 2000). This proteolytic processing was first described in the *Escherichia coli* phage HK97, where the precursor head protein is also cleaved on the C-terminal side of a lysine, which removes the first 103 amino acids to generate the mature capsid protein that begins with serine (Duda et al. 1995a, Duda et al. 1995b). It is believed this cleavage removes the portion of the capsid protein that acts as a scaffold for head assembly (Duda et al. 1995b). Based on the predicted protein sequence for *R. capsulatus*, this cleavage should occur separately with either the anti-GTA antibody or the anti-GTA antibody.
suppl/a059p283_app.pdf), and the R. capsulatus protein appears as a band of approximately 32 kDa on western blots (Fig. 3). All of the strains/species with positive results showed proteins of very similar size to the R. capsulatus protein, between 32 and 34.5 kDa (Fig. 3). An alignment of several capsid protein sequences from species within the Rhodobacterales shows this cleavage site is conserved across all these species, with either an alanine or a serine following the lysine residue (Fig. 4). Therefore, we predict that the other GTA capsid proteins, if produced, would also be processed at this site, and there is a strong congruence between the predicted sizes for mature GTA capsid proteins and the sizes of the proteins detected on the western blots (Table S2; Fig. 3), supporting this prediction. Conservation of this functionally important sequence is another piece of evidence indicating that these GTA genes in different strains/species are functional. Assaying capsid protein expression in more strains/species with complete genome sequences and determining the complete g5 sequences for the isolates with detected GTA proteins will allow this to be further confirmed.

Fig. 3. Expression of gene transfer agent (GTA) capsid proteins in phylogenetically diverse Rhodobacterales bacteria. (A) Expression of GTA capsid proteins in strains isolated from coastal Newfoundland. Top panel shows the results obtained with cells; bottom panel shows the results from the culture supernatants. (B) Expression of GTA capsid proteins in species/strains from other locations. Top panel shows the results obtained with cells; bottom panel shows the results from the culture supernatants. (C) Antibody neutralization experiment demonstrating the specificity of the anti-GTA capsid protein antibodies. Top panel shows the results obtained when the membrane was incubated with the anti-GTA capsid protein antibody; bottom panel shows the results when a duplicate membrane was incubated with the pre-neutralized antibody (+peptide); both panels contain samples of cells from the indicated strains/species. A side-by-side comparison of the full DMLB-w6 lanes is shown in Fig. S3, Supplement 1 (at www.int-res.com/articles/suppl/a059p283_app.pdf). Rc, Rhodobacter capsulatus SB1003; g5-, R. capsulatus A1; 4.2, DMLB-w4.2; 6, DMLB-w6; 8, DMLB-w8; 12.1, DMLB-w12.1; 13, DMLB-w13; DSS, Ruegeria pomeroyi DSS-3; Y4I, Phaeobacter sp. Y4I; SE, Citreicella sp. SE45; 05, CB1005; 23, CB1023; 40, CB1040. Sizes indicated on the left side correspond to molecular weight standards.

Fig. 4. Conservation of the Escherichia coli phage HK97 and Rhodobacter capsulatus RcGTA capsid protein-processing site in GTA protein sequences. Mature E. coli phage HK97 (Duda et al. 1995b) and R. capsulatus RcGTA (Lang & Beatty 2000) capsid proteins are proteolytically processed by cleavage on the C-terminal side of a lysine residue, indicated by the arrow. A protein alignment shows this cleavage site is conserved in the GTA capsid proteins from diverse Rhodobacterales species. Accession numbers for the sequences are provided in Table S1, Supplement 1 (at www.int-res.com/articles/suppl/a059p283_app.pdf).
SUMMARY

Despite the prevalence of members of the Rhodobacterales order in diverse marine systems, studies focusing on the biogeographical diversity of order members are limited. This is in part due to the lack of molecular tools that specifically target group members. The GTA g5 gene marker allows us to specifically target the Rhodobacterales community and provides high phylogenetic resolution for this important group of bacteria (Zhao et al. 2009). A remarkable diversity of the Rhodobacterales community was found in the sub-arctic water of the North Atlantic, suggesting that different bacterial species in this group are able to adapt to the cold environment. Our study shows that the genetic diversity of the Rhodobacterales community in the cold Subarctic ecosystem is higher than that in a temperate ecosystem, such as the Chesapeake Bay. The change in the bacterial community structure from warm to cold temperate regions also suggests specific adaptations of Rhodobacterales to different environments. The GTA gene cluster is a conserved genetic trait among this group of bacteria, the production of GTA has been demonstrated in 2 Rhodobacterales, and we have provided evidence that a number of additional strains from multiple locations express the GTA major capsid protein. It now becomes important to determine the rate or frequency of GTA-mediated genetic exchange in natural environments. The abundance of Rhodobacterales and the accumulating evidence of GTA production capacity by species in this order suggest GTA-mediated genetic exchange could be an important mechanism of homologous recombination in natural environments, particularly in marine systems where these bacteria can constitute a large proportion of the heterotrophic microbial community.

Acknowledgements. We thank K. Keats for assisting in sample collection, C. Way for the bacterial abundance data, C. Budinoff for providing the Chesapeake Bay isolates, J. T. Beatty for Rhodobacter capsulatus A1, and D. Campbell for advice on the antibody work. Y.F. and D.M. were supported in part by graduate fellowships from Memorial University of Newfoundland. Research in the laboratories of A.S.L. and R.B.R. was supported by grants from the Natural Sciences and Engineering Research Council (NSERC), the Canada Foundation for Innovation (CFI), and the Industrial Research and Innovation Fund (IRIF) from the Government of Newfoundland and Labrador. Research in the laboratories of A.B. and F.C. is supported by grants from the U.S. National Science Foundation.

LITERATURE CITED


Editorial responsibility: Jed Fuhrman, Los Angeles, California, USA

Submitted: September 1, 2009; Accepted: December 22, 2009
Proofs received from author(s): April 12, 2010