Abundant proteorhodopsin genes in the North Atlantic Ocean

Barbara J. Campbell, Lisa A. Waidner, Matthew T. Cottrell and David L. Kirchman*
College of Marine and Earth Studies, University of Delaware, Lewes, DE 19958, USA.

Summary

Proteorhodopsin (PR) is a light-driven proton pump that has been found in a variety of marine bacteria, including *Pelagibacter ubique*, a member of the ubiquitous SAR11 clade. The goals of this study were to explore the diversity of PR genes and to estimate their abundance in the North Atlantic Ocean using quantitative polymerase chain reaction (QPCR). We found that PR genes in the western portion of the Sargasso Sea could be grouped into 27 clusters, but five clades had the most sequences. Sets of specific QPCR primers were designed to examine the abundance of PR genes in the following four of the five clades: SAR11 (*P. ubique* and other SAR11 Alphaproteobacteria), BACRED17H8 (Alphaproteobacteria), HOT2C01 (Alphaproteobacteria) and an uncultured subgroup of the Flavobacteria. Two groups (SAR11 and HOT2C01) dominated PR gene abundance in oligotrophic waters, but were significantly less abundant in nutrient- and chlorophyll-rich waters. The other two groups (BACRED17H8 and Flavobacteria subgroup NASB) were less abundant in all waters. Together, these four PR gene types were found in 50% of all bacteria in the Sargasso Sea. We found a significant negative correlation between total PR gene abundance and nutrients and chlorophyll but no significant correlation with light intensity for three of the four PR types in the depth profiles north of the Sargasso Sea. Our data suggest that PR is common in the North Atlantic Ocean, especially in SAR11 bacteria and another marine alphaproteobacterial group (HOT2C01), and that these PR-bearing bacteria are most abundant in oligotrophic waters.

Introduction

Proteorhodopsin (PR) is a light-driven proton pump that may enhance ATP production by heterotrophic bacteria that otherwise rely on organic material oxidation for energy (Béjà et al., 2000; Martinez et al., 2007). Proteorhodopsin genes have been identified in a wide variety of cultured and uncultured bacterioplankton (Sabehi et al., 2003; 2004; 2005; de la Torre et al., 2003; Venter et al., 2004), including *Pelagibacter ubique*, a cultured representative of one of the most ubiquitous marine bacterioplankton clades, SAR11 (Giovannoni et al., 2005). In addition, PR is found in many Gammaproteobacteria, including members of the SAR86 and SAR92 clades (Sabehi et al., 2004; Stingl et al., 2007a). Recently, PR genes were identified in the genomes of cultured marine members of the Flavobacteriaceae, including *Polaribacter irgensii*, *Tenacibaculum* sp. MED152, *Dokdonia* sp. MED134 (Gomez-Consarnau et al., 2007) and a cultured representative of the gammaproteobacterial SAR92 clade, HTCC2207 (Stingl et al., 2007a). The phylogeny of the PR gene does not always coincide with those of typical phylogenetic marker genes, indicating that some bacteria and archaea may have obtained it via lateral gene transfer (Sabehi et al., 2004; Frigaard et al., 2006).

The abundance of PR genes has been estimated only indirectly from the frequency of PR-bearing clones in environmental genomic libraries. These estimates vary greatly, ranging from 13% of bacteria in surface water BAC libraries from the Mediterranean and Red Seas to 70% in the Sargasso Sea environmental genomic data set (Venter et al., 2004; Sabehi et al., 2005). In the recently published Global Ocean Survey (GOS), 2674 clones contained putative PR genes whereas 4125 clones contained 16S rRNA genes, suggesting that PR gene is very common (Rusch et al., 2007). Interestingly, 25% of the PRs clustered with the *P. ubique* PR gene, while the rest clustered with PR from several other bacterial types (Rusch et al., 2007). Beyond these estimates based on environmental genomic studies, there have been no quantitative assessments of PR gene abundance in oceanic waters.

Three major PR types are found in the oceans and these vary geographically and with depth apparently due to light quality (Béjà et al., 2001; Man et al., 2003; Rusch et al., 2007). Green light PR types are found in oceanic surface waters, whereas blue light types are more abundant in
deeper waters (Béjà et al., 2001). In addition, blue light-type PR genes dominate large insert clone libraries from oligotrophic waters and green light-type PR genes are more prominent in coastal environments (Rusch et al., 2007; Sabehi et al., 2007). This distribution of green and blue-type PRs correlates with the spectral qualities of these oceanic regimes (Rusch et al., 2007). About 20% of PR genes from both coastal and oligotrophic regimes belong to a third type, which may respond to the same wavelength as the green type of PR (Rusch et al., 2007).

The goals of this project were to estimate PR gene abundance in the North Atlantic Ocean and to examine how the abundances of different PR types vary with environmental parameters. Samples were collected during the 2005 North Atlantic Spring Bloom (NASB) cruise (Michelou et al., 2007). We used a real-time quantitative polymerase chain reaction (QPCR)-based approach to examine the abundance of four types of PR genes, including those closely related to PR in P. ubique. Prochlorophytochrome gene abundance data are needed to test hypotheses about what controls the frequency and abundance of PR-containing bacteria. We found that PR is abundant in marine bacterioplankton in the Sargasso Sea and in other areas of the North Atlantic Ocean and that the abundance of specific PR types varies with environmental conditions.

Results

Proteorhodopsin gene diversity

In order to identify PR gene groups for QPCR analysis, we generated a clone library of 91 PR gene sequences from a sample in the western Sargasso Sea (Station 1.5.1, 32.6°N, 65.6°W). The library contained 89 unique nucleic acid sequences and 76 unique amino acid sequences identified by DOTUR analysis (Schloss and Handelsman, 2005). Twenty-seven clusters were identified after grouping the sequences at 80% nucleic acid sequence similarity. Of these 27 clusters, five contained six or more clones each. Because of their overall amino acid similarity (72%), four clusters were combined to generate the BACRED17H8-like PR group. This group contained 26 clones from our PR library (Fig. 1).

All of the major clades in our clone library contained PR genes from either cultured representatives or phylogenetically identified large insert (BAC or fosmid) clones. Some of the topologies were not well resolved, based on the short sequence analysed (Fig. 1). Four clades contained sequences most similar to alphaproteobacterial PR gene sequences. Within the four groups identified as alphaproteobacterial-like, six PR clones were in the SAR11 clade containing P. ubique, with a mean amino acid similarity of 96%. The largest alphaproteobacterial clade (26 PR clones) contained sequences most closely related to the BACRED17H8 PR gene, with a mean amino acid similarity of 87%. The other two alphaproteobacterial groups were most closely related to the HOT2C01 fosmid PR gene (24 PR clones) and a PR gene from the genome sequence of the alphaproteobacterium HTCC2255 isolate (six PR clones), both with mean similarities of 94%.

Two clades did not contain alphaproteobacterial PR gene sequences. The first, containing six sequences, grouped with the PR sequence from MEDPR49C08, which is a BAC clone of the SAR86 group, a gammaproteobacterium (Béjà et al., 2000). This clade had a within-group mean similarity of 91%. The second major clade with nine PR sequences was in a larger diverse group containing three PR sequences from cultured Flavobacteria (Fig. 1) and had a within-group mean similarity of 97%. In addition to the major clades, there were several other PR sequences in the library. Most of these grouped with previously identified PR genes, many of which have not been phylogenetically characterized.

Estimates of PR gene abundance

Using the sequences of the PR genes amplified from the western Sargasso, we designed QPCR primers to target four groups of PR genes (Table 1): (i) the SAR11 group which contains P. ubique (Giovannoni et al., 2005), (ii) the group related to the fosmid PR sequence of HOT2C01, originating from an uncultured alphaproteobacterium (de la Torre et al., 2003), (iii) the group closely related to the PR sequence of the fosmid BACRED17H8, also most likely originating from an uncultured alphaproteobacterium (Sabehi et al., 2005), and (iv) the subgroup of NASB clones most closely related to the PR from the cultured flavobacterium, Dokdonia sp. MED134 (Gomez-Consarnau et al., 2007). We call this subgroup Flavo-NASB to emphasize that it does not include all types of PR from Flavobacteria. To check the specificity of the QPCR primers, we determined in silico which sequences in the GOS data set (Rusch et al., 2007) would be amplified by the set of primers for each group. The SAR11 PR primers would amplify 25% of the GOS PR sequences, with an overall mean amino acid similarity of 91%. Interestingly, this is the same percentage of PR sequences identified as SAR11 in the GOS data set (Rusch et al., 2007). The HOT2C01 and BACRED17H8 PR primers would amplify approximately 1% of the GOS PR sequences, with similarities of 95% and 89% respectively. None of the nearly full-length GOS
sequences matched both Flavo-NASB PR primers, although 7.5% matched the forward primer.

Abundances of PR genes in surface water samples from Ft. Pierce, Florida (79.5°W) to the Azores (29°W) were normalized to the abundance of bacterial 16S rRNA genes (Fig. 2). We chose to normalize PR gene copies to 1.9 copies of the 16S rRNA gene per bacterium because this is the ratio of 16S rRNA gene copies to the average of

**Table 1.** Primers and PCR conditions used in this study for examining 16S rRNA and proteorhodopsin (PR) genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Group</th>
<th>Region*</th>
<th>Primers (5′–3′)</th>
<th>Annealing temperature (°C)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR</td>
<td>SAR11</td>
<td>141–204</td>
<td>SARPR_125F SARPR_288R</td>
<td>54</td>
<td>This study</td>
</tr>
<tr>
<td>PR</td>
<td>BAC17H8</td>
<td>108–177</td>
<td>BAC_F2 BAC_R2</td>
<td>54</td>
<td>This study</td>
</tr>
<tr>
<td>PR</td>
<td>HOT2C01</td>
<td>116–189</td>
<td>HOT2C01_F HOT2C01_R</td>
<td>56</td>
<td>This study</td>
</tr>
<tr>
<td>PR</td>
<td>Flavo-NASB</td>
<td>103–146</td>
<td>Flavo_F2 Flavo_R2</td>
<td>56</td>
<td>This study</td>
</tr>
</tbody>
</table>

* Numbering based on *Escherichia coli* nucleotide sequence (16S rRNA) or *Pelagibacter ubique* translated amino acid sequence (PR).
six single copy genes in the Sargasso Sea (Venter et al., 2004).

Abundance of the bacterial group containing SAR11-like PR genes ranged from 4% to 24% of the total bacterial community along the entire transit from Ft. Pierce, Florida to the Azores (Fig. 2). The abundance of these SAR11-like PR bacteria was similar to the abundance of bacteria containing the HOT2C01-like PR gene (4–34% of the total community). The bacteria containing the BACRED17H8-like PR were less abundant (1.5–7%). The least abundant group was Flavo-NASB, a subgroup of clones most closely related to the PR of the cultured Flavobacteriaceae (0–3% of total bacterial community). Because the bacterial abundance estimated by 16S rRNA QPCR at 71°W (marked with a box in Fig. 2) was much less than expected, PR abundances calculated with this sample were not included in our statistical analyses, although inclusion of these values does not change our overall conclusions.

The SAR11-like and BACRED17H8 PR genes were approximately twofold more abundant in the middle and western parts of the Sargasso Sea (46–67°W) than in coastal waters and the eastern Sargasso Sea (Student’s t-test, \( P < 0.05, n = 16 \)) (Fig. 2). Proteorhodopsin genes from the HOT2C01 group were about fourfold more abundant in the Sargasso Sea between 67 and 61°W, and Flavo-NASB-like PR sequences were over 20-fold more abundant in the Sargasso Sea between 79 and 66°W than in areas outside these regions (Student’s t-test, \( P < 0.001, n = 16 \)) (Fig. 2).

To estimate total abundances of PR-containing bacteria, we added together the abundance estimates of all four tested microbial groups containing PR genes (Table 2). The total PR abundances ranged from <1% to 63% of total bacteria, assuming 1 PR and 1.9 16S rRNA gene copies per bacterium. We also averaged the PR gene abundances within the western Sargasso Sea and in other areas of the North Atlantic (Table 2). Except for the BACRED17H8-like PR, all of the PR gene types were more abundant in the western Sargasso Sea than in other areas. The number of total PR gene copies was

Table 2. Relative abundance of PR types in surface waters of the North Atlantic Ocean.

<table>
<thead>
<tr>
<th>PR type</th>
<th>Western Sargasso*</th>
<th>Otherb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Average</td>
</tr>
<tr>
<td>SAR11</td>
<td>0.14–0.24</td>
<td>0.18 (0.03)</td>
</tr>
<tr>
<td>BACRED17H8</td>
<td>0.04–0.05</td>
<td>0.05 (0.008)</td>
</tr>
<tr>
<td>HOT2C01</td>
<td>0.20–0.33</td>
<td>0.26 (0.04)</td>
</tr>
<tr>
<td>Flavo-NASB</td>
<td>0.004–0.03</td>
<td>0.02 (0.008)</td>
</tr>
<tr>
<td>Total</td>
<td>0.40–0.63</td>
<td>0.50</td>
</tr>
</tbody>
</table>

a. 69°W to 62°W, \( n = 3 \).
b. 79°W to 76°W, 61°W to 26°W, \( n = 18 \).

Proteorhodopsin gene copies were normalized to 1.9 16S rRNA gene copies (standard error in parentheses).
approximately twofold higher in the western Sargasso Sea than other areas in the North Atlantic (Table 2).

Variation in the abundance of PR genes with depth was examined at four stations along the Azores-Iceland transect (Fig. 3). Proteorhodopsin gene abundances in surface waters were similar to those in the Florida to Azores transit except for the HOT2C01-like PR, which was detectable only at the first two stations. In all cases, PR-containing bacteria decreased with depth. The largest decrease was observed at the first two stations closest to the Azores, where bacteria containing the SAR11 PR type decreased from approximately 10% in surface waters to 3% of the bacterial community at 200 m. Bacteria containing the HOT2C01 PR type at these stations decreased about two- to fourfold from the surface to 200 m. The SAR11 PR-containing bacteria at the second two stations decreased similarly (Fig. 3). While the SAR11 and HOT2C01 PR types were found at all depths, the BACRED17H8 and Flavo-NASB PR genes were less abundant, and were undetectable below 50–100 m.

**Relationship of PR-containing microbes to nutrients and chlorophyll a**

The variation in the abundance of PR-containing bacteria appeared to be due to nutrient and chlorophyll a concentrations and not light in the depth profiles (Table 3). Relative gene abundance of three of the four PR types and of the sum of all PR types decreased significantly (negative correlation) with phosphate, silicate (except for HOT2C01), and nitrate + nitrite concentrations in the depth profiles (Table 3). The notable exception was the lack of any significant correlation between Flavo-NASB PR abundance and these nutrients. Even for the significant correlations, the relationship between PR gene abundance and nutrient concentrations such as phosphate was non-linear (Fig. 4A) as was the relationship with chlorophyll (Fig. 4B). The relative abundances of two PR types (SAR11 and HOT2C01) and of the sum of all four PR genes measured in this study were negatively correlated with chlorophyll a concentrations in the depth profiles (Fig. 4B;

<table>
<thead>
<tr>
<th>PR type</th>
<th>Chlorophyll a</th>
<th>Phosphate</th>
<th>Silicate</th>
<th>Ammonium</th>
<th>Nitrate</th>
<th>Light</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR11</td>
<td>–0.767**</td>
<td>–0.773**</td>
<td>–0.600**</td>
<td>–0.327</td>
<td>–0.733**</td>
<td>0.501*</td>
</tr>
<tr>
<td>BACRED17H8</td>
<td>–0.515</td>
<td>–0.640**</td>
<td>–0.471*</td>
<td>–0.323</td>
<td>–0.572**</td>
<td>0.357</td>
</tr>
<tr>
<td>HOT2C01</td>
<td>–0.642*</td>
<td>–0.578**</td>
<td>–0.427</td>
<td>–0.426</td>
<td>–0.486*</td>
<td>0.106</td>
</tr>
<tr>
<td>Flavo-NASB</td>
<td>0.049</td>
<td>–0.166</td>
<td>–0.414</td>
<td>0.038</td>
<td>–0.204</td>
<td>0.243</td>
</tr>
<tr>
<td>All</td>
<td>–0.744*</td>
<td>–0.722**</td>
<td>–0.562*</td>
<td>–0.394</td>
<td>–0.652**</td>
<td>0.336</td>
</tr>
</tbody>
</table>

*Correlation is significant (P < 0.05). **Correlation is significant (P < 0.01).

The gene abundance of each PR type was normalized to 16S rRNA gene copies. The number of samples was 20, except there were 10 samples for the chlorophyll analyses. ‘All’ refers to the correlation analyses of the sum of all four PR types versus the indicated parameters.

© 2007 The Authors
Table 3). In surface samples only, two PR types and the sum of all four decreased significantly with chlorophyll concentrations ($r = -0.713$, $-0.614$ and $-0.671$ for BACRED17H8, HOT2C01 and the sum respectively; $P < 0.02$; $n = 15$) while the relationship was insignificant for Flavo-NASB ($r = -0.110$; $P > 0.05$; $n = 15$) and barely so for SAR11 PR ($r = -0.508$, $P = 0.053$; $n = 15$). Overall, PR gene abundance was highest when chlorophyll a was $< 1 \mu g \text{ liter}^{-1}$.

In contrast to nutrients and chlorophyll, there was a significant correlation between the relative abundance of only one of the four PR types (SAR11) and relative light intensity (% of surface irradiance) in the depth profiles (Table 3). There was no significant correlation between light and the other three PR types and with the sum of all four ($r < 0.40$; $P > 0.05$; Table 3).

Abundance of SAR11-like PR and SAR11 bacteria

Abundance of SAR11 PR-containing bacteria and total abundance of all SAR11 bacteria followed similar trends along the Ft. Pierce, Florida to Azores transit (Fig. 5). SAR11 bacterial abundance in the North Atlantic Ocean was measured by QPCR with primers specific to the SAR11 clade (Suzuki et al., 2001) and normalized to all bacteria using the general 16S rRNA QPCR primer pair (BACT1). SAR11 16S abundance was similar to the abundance of the SAR11-like PR gene and ranged from 4% to 28% of all bacteria (Fig. 5). The SAR11 PR and SAR11 16S rRNA gene estimates were positively correlated for all samples ($r = 0.84$; $P < 0.001$; $n = 37$). SAR11 bacteria containing the SAR11 PR gene type averaged 81%, and ranged from 37% to 120%. Two samples in the western Sargasso Sea had higher than expected abundances of the PR gene, resulting in lower ratios (Fig. 5). When the Azores to Iceland samples were included in the analysis, the average increased to 106% with a standard error of 9%.

The abundance of SAR11 bacteria estimated by QPCR was about twofold lower than that estimated by fluorescence in situ hybridization (FISH) (Fig. 5). In one station (57°W), SAR11 abundance as measured by FISH was fourfold higher than QPCR values of the SAR11-like PR gene and the SAR11 16S rDNA gene. However, SAR11-like PR gene copies were significantly correlated with SAR11 cell abundance as measured by FISH in samples from 79 to 70°W ($r = 0.97$, $P < 0.01$, $n = 3$) and 51 to 35°W ($r = 0.78$, $P < 0.01$, $n = 6$). The relationship between bacteria containing the PR gene and cells detected with the SAR11 FISH probe set was not significant when all data were analysed together ($r = 0.29$, $P > 0.05$, $n = 16$).

Discussion

Several studies have examined the diversity and geographic distribution of PR genes, but there are few estimates of PR gene abundance in the oceans. Unlike other prokaryotic phototrophs (e.g. cyanobacteria and aerobic...
Anoxygenic phototrophic bacteria, PR-containing microbes cannot be quantified using direct methods such as microscopy or flow cytometry (Schwalbach and Fuhrman, 2005; Cottrell et al., 2006; Michelou et al., 2007). Our study provides the first abundance estimates of several PR types in oceanic waters using QPCR and allows us to begin to use these data to understand what controls PR distribution in the oceans. The data from our study suggest that PR genes are present, on average, in at least 50% of marine bacteria in the western Sargasso Sea.

Proteorhodopsin gene diversity and abundance

Our results are similar to the estimates based on the shotgun sequence data sets from the Sargasso Sea, where at least 65% of the bacterial community contained a PR gene (Venter et al., 2004; Rusch et al., 2007). These estimates are much higher than the 13% estimate for the euphotic zone of the Mediterranean Sea based on a large insert (BAC) library approach (Sabehi et al., 2005). The difference between these estimates is surprising, because these two environments are both nutrient-poor regimes. The difference in PR estimates could be due to spatial and temporal variation in PR-containing bacteria, such as we observed here. Also, the many methodological differences between constructing small (Venter et al., 2004; Rusch et al., 2007) and large (Sabehi et al., 2005) insert libraries may lead to different estimates of relative PR abundance. In particular, large inserts bearing SAR11 PR genes may not be easily cloned into BAC or fosmid vectors. For instance, SAR11-like PR containing clones were not found in over 10,000 clones from two BAC libraries from the Mediterranean and Red Seas (Sabehi et al., 2005) or from over 5000 fosmid clones from a depth profile study of Station ALOHA samples in the North Pacific Gyre (DeLong et al., 2006), even though the SAR11 clade is an abundant, cosmopolitan clade (Field et al., 1997; Morris et al., 2002; Pommier et al., 2007).

The PR genes examined by this study appear to be common in the oceans and are not restricted to the regimes we examined. In all cases except one (NASB PR clone 28), the PR genes identified in this study were at least 70% similar at the amino acid level to PR genes found by previous studies. Based on similarities to either cultured representatives or to large insert clones with a characteristic phylogenetic marker, at least three major groups were within the Alphaproteobacteria, one within the Gammaproteobacteria and one within the Flavobacteria. Most of our PR clones were very similar to PR sequences in known Alphaproteobacteria and Flavobacteria clades. We chose to quantify the PR gene from these groups because of their presence in a wide range of ocean regimes (Pommier et al., 2007), enabling future QPCR-based studies to use these PR primers in other marine waters. Our analyses suggest that our primers would retrieve 1–25% of the PR sequences found by the GOS study, which sampled surface waters ranging from eutrophic coastal regimes to the oligotropic Pacific Ocean (Rusch et al., 2007). We did not target the gammaproteobacterial SAR86 bacteria group in our QPCR assay because of their low abundance in the Sargasso Sea shotgun library (Venter et al., 2004) and in our PR clone library from the western Sargasso Sea.

According to the quantitative abundance data, PR genes most closely related to the alphaproteobacterial clone HOT2C01 and P. ubique were most abundant in the Sargasso Sea and just north of the Azores. The original HOT2C01 PR BAC clone was found by screening a library generated from surface bacterioplankton at Station ALOHA in the North Pacific Gyre (de la Torre et al., 2003). In order to estimate abundances of particular subgroups, we chose sequences in the Sargasso Sea shotgun library that contained at least 60% of the entire PR gene (Venter et al., 2004). Of those, approximately 43% and 5% of the PR sequences were in the SAR11 and HOT2C01-like clades respectively. The third most abundant PR type identified in the NASB samples was most closely related to the PR gene in the BACRED17H8 large insert clone from a 7 m Red Sea BAC library (Sabehi et al., 2005). The BACRED17H8 PR made up 5% of the PR genes in a Mediterranean Sea BAC library (Sabehi et al., 2005). This PR group was also found in the Sargasso Sea and ALOHA metagenomic data sets, with percentages of 4% and 11% respectively (Venter et al., 2004; DeLong et al., 2006).

These results suggest that the three most abundant PR types that we quantified are common in oligotrophic regions of the North Atlantic and North Pacific Ocean and the Mediterranean Sea.

The total abundance of the four quantified PR genes was the highest in the Sargasso Sea at 66°W where the sample for the PR gene clone library was taken. This is not surprising, as we selected the PR genes to quantify based on the sequences obtained from that library. However, estimates of abundance based on their frequency in the clone library were not consistent with their abundance as estimated by QPCR. For instance, based on the PR library, we would have estimated that the SAR11 PR gene type would make up 9.2% of the total. In contrast, our QPCR data indicated that this group comprised 36% of the combined four PR gene types. The two other groups (BACRED17H8 and Flavobacteria) were somewhat overestimated in the clone library, while the HOT2C01 PR types were somewhat underestimated. For example, the HOT2C01 type comprised 37% in the library versus 52% with the QPCR approach. Our results indicate that estimates of gene abundance based on frequencies in clone libraries need to be cautiously interpreted.
SAR11 16S rRNA gene and SAR11-like PR gene abundances were significantly correlated and, based on our average estimates, an equal number of SAR11 bacteria detected by QPCR also contained a PR gene. These results are similar to those from a study of cultured SAR11 bacteria from the Atlantic and Pacific Oceans (Stingl et al., 2007b). The abundance of SAR11 estimated with both of these QPCR primer sets was approximately half that of the FISH analysis, which used a four-probe mixture. It is likely that the primers used in our study did not retrieve all SAR11-like 16S rRNA or PR genes present in the samples. For instance, sequences of 16S rRNA genes in the SAR11 group can be up to 15% dissimilar (Field et al., 1997; Rusch et al., 2007), and therefore, we could have missed some SAR11 bacteria with this 16S rRNA gene primer set. Another potential reason for our underestimate based on QPCR is the assumption of one PR copy and 1.9 copies of the 16S rRNA per bacterial genome. If we assume one copy of the 16S rRNA gene per bacterium, which is the case for the cultured representative of the SAR11 clade (Giovannoni et al., 2005), then the number of SAR11 bacteria estimated by QPCR would be closer to, but still less than the FISH-based estimate. Alternatively, the FISH probes could bind to closely related groups outside of the SAR11 clade, leading to overestimates of SAR11 bacterial abundance. It would be difficult to resolve this issue, unless techniques such as a combined single-cell PCR-based approach were used (e.g. Ottesen et al., 2006).

Environmental controls of PR gene abundance

We observed that the gene abundance of three of four PR types decreased with increasing nutrient and chlorophyll concentrations. Low nutrients could directly select for PR-containing bacteria because the extra energy gained by phototrophy could facilitate transport of phosphate and nitrate occurring in low concentrations; phosphate may limit heterotrophic bacterial growth in some of these waters (Cotner et al., 1997; Obemosterer et al., 2003). In addition, both low nutrient and low chlorophyll concentrations can directly or indirectly lead to a low supply of DOM, which also has been shown to affect bacterial growth in regimes like the Sargasso Sea (Carlson et al., 2002). In low-DOM environments, heterotrophic bacteria that can obtain extra energy from light could have an advantage over those without light-harvesting capabilities (Karl, 2002). Note, however, that the four PR types examined here did not vary the same in response to light, nutrients and chlorophyll, and one type (Flavo-NASB) did not correlate at all with these environmental parameters. There results are consistent with the GOS data set, where different PR gene types are found in open and coastal oceanic waters (Rusch et al., 2007).

It was surprising that PR gene abundance did not correlate more strongly with relative light intensity in our depth profiles, but light did have a significant effect on the SAR11 PR gene abundance and probably has some role in determining the distribution of the other PR types we observed. Although a number of amino acids contribute to PR binding to retinal (Bielawski et al., 2004), the amino acid found at position 105 (based on eBAC31A08 numbering) is most responsible for PRs responding to either blue (490 nm) or green (530 nm) light (Man et al., 2003). Almost all NASB PR clones from our diversity study contained a glutamine (blue-type amino acid) at amino acid position 105. In particular, all of the SAR11-like PR genes in our diversity study and in the Sargasso Sea shotgun library data set (Venter et al., 2004) are the blue type. Small insert clone libraries from oligotrophic and coastal surface waters are dominated by blue-type and green-type SAR11 PR gene variants respectively (Rusch et al., 2007; Sabehi et al., 2007). Unfortunately we cannot use our QPCR data to explore more extensively how variation in spectral tuning explains variation in SAR11 and other PR-bearing bacteria because our QPCR primers did not flank the nucleotide positions encoding amino acid 105.

Conclusions

Our data suggest that PR-containing bacteria, in particular those in the SAR11 clade, are abundant in the North Atlantic Ocean, accounting for at least 50% of marine bacterioplankton in the oligotrophic Sargasso Sea. However, some marine waters may be dominated by other types, including those in the alphaproteobacterial PR clade represented by the BAC clone HOT2C01, found in other oligotrophic oceans (de la Torre et al., 2003; DeLong et al., 2006). In addition, our study further supports the notion that different PR types are abundant in different ocean regimes (Rusch et al., 2007) and indicates that factors such as nutrient and DOM concentrations in addition to light may control the abundance of different types of PR in oceanic waters. Our PR gene abundance data support the hypothesis that bacteria containing PR and perhaps other photoheterotrophs are likely to be more successful in oligotrophic waters. Additional experiments are clearly needed to characterize the abundance, expression and regulation of PR genes in other oceanic regimes and in cultured marine prokaryotes.

Experimental procedures

Sampling and environmental parameters

Water samples were collected during the NASB project on board the R/V Seward Johnson in May–June, 2005. Surface
seawater for sampling was collected daily from the ship’s underway system from about 5 m depth from Ft. Pierce, Florida to Ponta Delgada, Azores. Seawater from various depths in the photic zone was collected using a rosette of Niskin bottles mounted on a CTD profiler from the Azores to Iceland (Michelou et al., 2007). Photosynthetically active radiation was measured at various depths with a sensor mounted on the profiler. The light attenuation coefficient was calculated by linear regression of log-transformed irradiance versus depth and used to calculate the percentage of surface irradiance at each sampling depth. These relative light intensities at each depth were subjected to the arcsin transformation before statistical analyses. Concentrations of NO$_3$ + NO$_2$, PO$_4$ and NH$_4$ were determined by automated, segmented flow colorimetric analysis on board ship. Chlorophyll a was measured fluorometrically using GF/F (Whatman) filters extracted overnight in 90% acetone (Parsons et al., 1984).

For clone library and QPCR analyses, all water samples were pre-filtered through 0.8 µm polycarbonate filters to minimize eukaryotic DNA. The bacterial size fraction was collected on 0.22 µm Durapore filters, placed in sucrose lysis buffer, and stored frozen in Whirlpak bags at −80°C. DNA was extracted using standard protocols (Fuhrman et al., 1988) and quantified via spectrophotometry. For QPCR, DNA was further purified using the Isoquick Nucleic acid extraction kit according to the manufacturer’s instructions (ISC Bioexpress, UT). DNA concentrations were quantified via a standard picogreen assay (Invitrogen, Carlsbad, CA) on a POLARstar Optima fluorometer (BMG Labtech).

Abundance of the SAR11 clade in the surface water samples from the Florida to Azores transit was determined using FISH (Malmstrom et al., 2005). Briefly, a SAR11-specific probe mixture of four 16S rRNA gene oligonucleotides was used to identify SAR11 bacteria (Morris et al., 2002) at 42°C in a hybridization buffer containing 30% formamide. Non-specific binding was estimated in hybridizations with a negative control probe (Karan and Fuhrman, 1997) and averaged 5.4% of all cells.

**Diversity analysis of PR genes**

Diversity of PR was assessed using degenerate PR gene primers (Sabehi et al., 2005). For this work the PCR primers were slightly modified with a nonsense extension at the 5′ end to improve amplification: PR-1aF: 5′-GATCGACCGNTAYRTHGARTGG-3′; PR-1aR: 5′-GATCGACGATADATNGC CCANCC-3′. Ten nanograms of DNA from Station 1.5,1 surface water (32°59.98′N, 65°59.85′W) was used in a PCR amplification, with the following conditions: 1 cycle of 94°C, 2 min, followed by 30 cycles of 94°C, 30 s; 52°C, 30 s; and 72°C, 30 s. Final concentrations of primers and MgCl$_2$ were 1.87 µM and 2 mM respectively. The PCR product was subjected to electrophoresis through 1.5% agarose, and a product of the expected size (approximately 335 bp) was excised from the gel and eluted with a Gene Clean gel extraction kit (Q-Biogene).

The purified fragment was cloned into a TOPO TA cloning kit for sequencing (Invitrogen) according to the manufacturer’s instructions. Colonies were screened via PCR to determine the insert size and 96 clones of the expected size were sequenced by standard Sanger sequencing. Sequences were trimmed of vector sequence and aligned using the CLUSTAL program in Mega version 3.1 (Kumar et al., 2004). Representative sequences most closely related to the NASB PR clone sequences were retrieved from GenBank and added to the alignment. Within-group mean distances were calculated in Mega 3.1. Phylogenetic trees of the translated sequences were constructed in Mega 3.1 with the default parameters, based on both neighbour-joining and minimum evolutionary algorithms with 500 bootstrap replicates. Proteorhodopsin gene sequences from the clone library and QPCR specificity test were submitted to GenBank and assigned the Accession No. EU079138–EU079274.

**Quantification of PR gene abundance**

To determine the numbers of PR genes in the NASB samples, four PR gene groups were targeted for QPCR (Table 1). Proteorhodopsin primers were designed from the nucleotide sequences of our NASB clones and representative sequences from GenBank, including over 100 from the Sargasso Sea data set (Venter et al., 2004). In addition, the 16S rRNA genes for all bacteria and for the SAR11 group were quantified using primers and conditions described elsewhere (Suzuki et al., 2000).

Four plasmid clones from our Sargasso PR library were used as positive controls and in standard curves in the QPCR assays. In addition, a clone containing a SAR11 16S rRNA gene (OSCI43, U75266) (Rappé et al., 2000) was used for a control for estimating the abundance specifically the 16S rDNA gene from SAR11 and of all 16S rDNA genes. Plasmids were isolated, linearized with PstI and isolated as described above. Products were quantified in triplicate using the picogreen assay. Standard reactions containing linearized plasmid contained approximately 10$^7–10^8$ copies per reaction. All standard curves were linear within the ranges tested. DNA from the indicated sample was diluted to approximately 100 pg µl$^{-1}$ and quantified again using the picogreen assay. Table 1 lists the gene, region, primers, primer concentrations and annealing temperatures for QPCR.

Quantitative PCR was performed in triplicate or quadruplicate with 1 µl of diluted DNA in a final volume of 12.5 µl using the Stratagene SYBR green mix on the ABI 7500, and PCR conditions of (Waidner and Kirchman, 2007): 95°C 10 min; followed by 30–40 cycles of amplification at 95°C for 15 s, the indicated annealing temperature for 45 s, and 72°C for 45 s, with a final dissociation step. Final primer concentrations were 0.2 µM except for the general 16S rRNA primers (0.096 µM) and the Flavo-NASB PR primer set (0.16 µM). Only single peaks were observed in the dissociation curves from both the standards and samples, indicating specific amplification with each set of primers. Average amplification efficiencies were as follows: general 16S rDNA = 87%; SAR11 16S rDNA = 99%; SAR11 PR = 77%; HOT2C01 PR = 65%; BAC17H8 PR = 66%; and Flavo-NASB PR = 97%.

To verify primer specificity, the four QPCR primer pairs were used in PCR to amplify PR genes from the Station 1.5.1 sample. Polymerase chain reaction amplification conditions were as outlined above and in Table 1, except no dissociation step was used. Polymerase chain reaction products were cloned and sequenced as described above.

© 2007 The Authors
Journal compilation © 2007 Society for Applied Microbiology and Blackwell Publishing Ltd, Environmental Microbiology, 10, 99–109
were analysed from each primer set. The sequences of the amplification products were as expected in all cases except with the SAR11 PR primer set, in which 11 out of 12 PR gene sequences (92%) were SAR11-like, and one clone was most closely related to a PR from a Flavobacterium (data not shown).

The estimates of PR gene abundance were normalized to 16S rRNA gene abundance in order to estimate the fraction of the bacterial community bearing each PR gene type. We assumed one PR copy and 1.9 16S rRNA copies per genome. The latter assumption is the average (SD = 0.39) ratio of 16S rRNA gene copies to six single-copy genes (rpoB, recA, atpD, gyrB hsp70 and tufA) in the Sargasso Sea shotgun library (Venter et al., 2004). Errors on estimates of relative PR gene abundance were calculated from the errors in estimating PR and 16S rRNA gene abundance and standard propagation of error equations.

Acknowledgements

We wish to thank D.A. Hutchins, G.M. Berg and S.W. Wilhelm for supplying nutrient and chlorophyll data from the NASB cruise, V. Michelou for sampling assistance and L. Yu for technical assistance. The SAR11 16S rRNA gene clone was supplied by M. Rappé. Support for this project was provided by grants from the National Science Foundation.

References


