ABSTRACT

Title of Thesis:	MEASUREMENT OF *IN SITU* EXPRESSION OF PROTEORHODOPSIN GENES AT THE NORTH PACIFIC CENTRAL GYRE STATION ALOHA
	Yanmei Shi, Master of Science, 2005
Thesis Directed By:	Dr. Marcelino T. Suzuki Marine Estuarine Environmental Science

A novel approach was developed for the measurement of microbial gene expression in the oceans, and used to test the hypothesis that Proteorhodopsin (PR) gene expression is regulated at the transcriptional level. Chapter 1 describes the optimization of total nucleic-acid extraction that allows "quasi-absolute" quantification of PR mRNAs and coding genes, and the subsequent use of mRNA to gene ratio as a measure of PR gene transcription. In Chapter 2, I describe the construction of a PR cDNA clone library using an RNA sample collected from the North Pacific Central Gyre Station ALOHA, from which diverse PR gene sequences were retrieved and subjected to phylogenetic analyses based on the amino acid sequences. Chapter 3 describes the measurement of *in situ* PR gene transcription during three diel cycles. Light availability and physical processes such as advection were both found to affect measured PR gene transcription at 25 m at station ALOHA.

MEASUREMENT OF *IN SITU* EXPRESSION OF PROTEORHODOPSIN GENES

AT THE NORTH PACIFIC CENTRAL GYRE STATION ALOHA

By

Yanmei Shi

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2005

Advisory Committee:

  Assistant Professor Marcelino T. Suzuki, Chair
  Professor Diane K. Stoecker
  Assistant Professor Feng Chen

ACKNOWLEDGMENTS

Finally, I am especially grateful to my parents and other family members, for without their love, support and encouragement, this work could never be possible and I could never be where I am right now.

Table of Contents

List of Tables

List of Figures

Introduction to Proteorhodopsins and Microbial Gene

Expression in the Ocean

Bacterioplankton account for a large proportion of total planktonic biomass, and perform essential roles in marine ecosystems, including primary production, nutrient cycling and energy flux (Azam 1998; Azam and Worden 2004; Béjà et al. 2002). However, microbial ecologists have long noted that there is a great discrepancy between total microbial cell counts obtained via microscopic observation from colony numbers on agar plates, which was named the "Great plate count anomaly" (Staley and Konopka 1985). Subsequently, numerous studies showed that most microbes in the oceans are viable in the environment but refractory to conventional cultivation techniques (Giovannoni and Rappé 2000; Pace 1997), which in turn precludes the full understanding of microbial diversity and function in the environment. Phylogenetic studies of environmental 16S rRNA genes have fundamentally changed earlier paradigms of bacterioplankton community structure and a large diversity of unique and previously unrecognized microorganisms is emerging, with their physiology and functions awaiting further exploration. Additionally, the introduction of environmental genomics (DeLong 2004; Rodriguez-Valera 2004) and "shotgun" sequencing approaches (Tyson et al. 2004; Venter et al. 2004) has further resulted in the discovery of millions of novel genes that encode for proteins with diverse functions, which confirmed the fact that much of the microbial functional capabilities and controlling mechanisms are unknown.

Proteorhodopsins

In a recent study by Ed Delong's group (Béjà et al. 2000a), a new type of phototrophy in the ocean was indicated by the discovery of a proteorhodopsin (PR)-coding gene in one uncultured *Gammaproteobacterium* from the SAR86 clade

inhabiting surface waters of the Monterey Bay, California. Proteorhodopsins are retinal-binding membrane proteins that share high amino acid sequence similarity with archaeal bacteriorhodopsins (BRs), and can function as proton pumps to convert light energy to cellular energy in the form of synthesized ATP (Béjà et al. 2000a). Subsequent studies indicated that a wide variety of PR genes with different spectral properties are distributed throughout marine photic zones in accordance to light quality and availability (Béjà et al. 2001; Man-Aharonovich et al. 2004; Sabehi et al. 2003), and are harbored by divergent bacterial taxa including *Alpha-* and *Gammaproteobacteria* (de la Torre et al. 2003; Sabehi et al. 2004). Furthermore, a massive shotgun cloning and sequencing effort in the Sargasso Sea has resulted in the discovery of over 600 novel rhodopsin homologs belonging to 13 distinct rhodopsin sub-families (Venter et al. 2004). This finding dramatically changed existing notions about the abundance and diversity of naturally occurring PRs, suggesting high abundance, diversity, and ubiquitousness of PR-bearing bacteria in the marine waters. Although it is unlikely that PR-bearing bacteria can fix CO_2 and contribute to primary production as phytoplankton and to some extent, aerobic anoxygenic phototrophs do, their potential significant contribution to energy fluxes in the oceans may revise previous views regarding oceanic phototrophy.

Due to this potential ecological significance of PRs, there is a growing interest on the exploration of PR gene expression in the oceans. However, despite a large collection of PR-related sequences obtained from various ocean sites either by PCR-based or genomics-based methods (Béjà et al. 2001; Sabehi et al. 2003; Venter et al. 2004), not much is known about PR expression *in situ*. Stapels and coworkers

(Stapels et al. 2004) successfully extracted and identified proteorhodopsins from a *Alphaproteobacterium* isolate from the Pacific Ocean, which indicated that proteorhodopsin is expressed in this cultivated strain and thus, likely *in situ.* This strain has been classified as *Pelagibacter ubique* and belongs to the ubiquitous and abundant SAR11 clade (S.J. Giovannoni, personal communication). In addition, biophysical analyses of membrane preparations of bacterioplankton directly collected from Monterey Bay surface waters suggested the physical existence and functional activity of proteorhodopsins in the Eastern Pacific (Béjà et al. 2001). Although *in situ* protein profiles are the most straightforward and robust evidence for PR gene expression in the environment, the purification and quantification of specific PR components from complex protein mixtures remains very challenging.

Alternatively, the currently available PR gene sequence information allows the exploration of *in situ* PR mRNA diversity by reverse transcription (RT) PCR with degenerate primers, providing a new avenue for the understanding of the identity and diversity of the actively expressed PR genes in the oceans. In chapter 2, we describe the results of a phylogenetic analysis of PR cDNAs retrieved from station ALOHA in the North Pacific Gyre along with all other known PRs, concluding that diverse bacterial groups (though, mostly belonging to the *Alphaproteobacteria*) harbor and express PR genes in the North Pacific photic zone.

<u>Measuring gene expression in the ocean</u>

In addition to identifying active members of the PR-containing bacterial community, mRNA molecules can be detected with a high degree of specificity using real-time RT-PCR techniques and provide a real-time reflection of gene expression

due to their relatively short half-life. Thus, mRNA levels have been broadly used in gene expression studies as surrogates for the expression levels of transcriptionally regulated genes (Wawrik et al. 2002; Wilson et al. 1999). Genes that are regulated at least in part at the transcriptional level, such as *rbc*L (Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase) and *gln*A (Glutamine Synthetase), show rhythms of mRNA abundance corresponding to regulators of gene expression in the environment (Wawrik et al. 2002; Wyman 1999). Though little is known about the regulation mechanisms of PR genes due to the difficulty associated with the cultivation of PR-containing bacteria, current evidence points to transcriptional level regulation (Baliga et al. 2001). Thus, understanding how PR mRNA levels respond to the variation of possible PR gene expression regulators in the environment may provide further insights into the regulation of PR genes.

Several potential factors may cause changes in mRNA levels that could be erroneously interpreted as variation in gene expression. Nucleic acid extraction efficiency might differ among samples, resulting in different amounts of starting material for quantification (Weinbauer et al. 2002). Reverse transcription reaction efficiency also may vary among samples and lead to errors (Ståhlberg et al. 2004b). Finally, changes in mRNA signals in naturally occurring populations might also be caused by changes of cell numbers (gene abundance, Pichard and Paul 1993; Wyman 1999), or changes of bacterial community structure due to physical processes such as advection, rather than transcriptional activity. Therefore, in order to obtain biologically relevant transcriptional activities for target genes in environmental samples, accurate quantification and proper normalization of target mRNA levels is

critical. To date, most gene expression studies quantifies target mRNA molecules relative to mRNA levels of certain calibrator genes (e.g., rRNA genes, housekeeping genes), cell counts, or total RNA amount (Bustin 2000). However, these normalization approaches are problematic for environmental RNA samples extracted from highly heterogeneous microbial populations because: 1) normalization of target mRNA to calibrator gene mRNA or cell counts might be hindered by the lack of knowledge regarding the linkage between target genes and the corresponding housekeeping gene, rRNA gene and cell counts, and 2) normalization of target mRNA to total RNA might lead to biased results due to potential effects of changes in target gene abundance and bacterial community structure.

Alternatively, the normalization of target mRNA abundance to target gene abundance does not require previous knowledge of the linkage genes, since target mRNA and target gene are directly linked, and can compensate for the variation in target organism abundance, allowing the measurement of biologically relevant expression levels. However, the use of mRNA-over-gene ratio as measurement of gene expression is only valid with unbiased quantification of the target gene and the associated mRNA. In most previous studies focusing on environmental microbes, target genes (such as the 16S rRNA gene), and target mRNAs were quantified relative to community-wide parameters such as total bacterial 16S rRNA genes and total RNA amounts, respectively. However, these approaches face the risk of data misinterpretation due to possible microbial community change, since a seemingly increase (or decrease) of target gene (or mRNA) signal might result from decreases (or increases) of non-target gene (or mRNA) abundance. Thus, absolute

quantification of target gene and mRNA in the environment (as copy number per volume seawater) would be highly preferable.

In theory, the high specificity and accuracy of real-time quantitative PCR techniques could allow absolute quantification of target genes and mRNAs in the environmental samples. In practice, efficiencies of nucleic acid extraction and reverse transcription are unlikely to reach 100%. However, as long as they are constant among samples, quantification of target genes and mRNAs in a "quasi-absolute" fashion and among-sample comparisons are possible, circumventing problems associated with relative quantification. Chapter 1 of this thesis describes the optimization of RNA and DNA extraction methods and the evaluation of the constancy of RNA extraction efficiencies among samples, and the development of a method that enables quantification of PR mRNAs and genes in a "quasi-absolute" fashion. Finally, in Chapter 3 we applied this method to quantify PR genes and mRNAs and measure biologically relevant *in situ* PR gene expression, using the ratio between PR mRNA and PR gene as an indicator for PR gene expression levels and explored possible effects of physical variables such as light and advection on PR gene expression levels *in situ*.

CHAPTER 1: Optimization of Total RNA and Proteorhodopsin mRNA Extraction

Abstract

Absolute quantification of specific mRNAs in environmental samples (as number of copies per sample volume or mass) is dependent on the assumption that mRNA extraction efficiency is constant among different samples. Here, we report on the development of an optimized protocol for RNA extraction from bacterioplankton, based on the RNeasy[®] protocol (Qiagen, Valencia, CA) that allows "quasi-absolute" measurements of mRNA levels in marine and estuarine bacterioplankton samples. Seeking to maximize RNA yields of the extraction, five different types of beads used for cell disruption were compared and low protein binding Zirconium beads provided the highest total RNA yields. Subsequently, we used this optimized extraction protocol to evaluate the constancy of RNA extraction efficiency (total RNA and proteorhodopsin mRNA) among samples, by comparing sub-samples filtered from increasing volumes of a water sample from the North Pacific Central Gyre station ALOHA. Preliminary experiments with bacterioplankton samples from the Patuxent Estuary suggested that the upper limit for linear elution of RNA from Qiagen RNeasy[®] spin columns might have been reached for sample volumes above 400 ml. On the other hand, although station ALOHA bacterioplankton samples were filtered from water volumes as large as 5 l, extracted total RNA did not appear to reach this upper limit, likely because of the significantly lower RNA content in open ocean waters. Using the optimized protocol, both total RNA and proteorhodopsin (PR) mRNA concentrations in the RNA extract showed linear relationships with subsample volumes (R^2=0.93 and 0.94, respectively). Total RNA extraction efficiency was not significantly different in the 1, 2, 4, and 5 l subsamples (ANOVA, p=0.34), but the

0.5 l sample had significantly higher extraction efficiency in station ALOHA bacterioplankton samples. PR mRNA extraction efficiency was not significantly different for 4 and 5 l subsamples (ANOVA, p=0.37), but the remaining subsamples apparently had lower extraction efficiencies. However, since PR mRNA was reverse-transcribed to cDNA before quantification, the apparent lower extraction efficiency of PR mRNA more likely reflected lower reverse transcription rates at lower mRNA concentrations.

Introduction

The recent application of nucleic acid-based molecular techniques has circumvented difficulties associated with the cultivation of naturally occurring marine microorganisms, and has greatly advanced our knowledge of the identity and function of microbes in the environment (Béjà et al. 2000a; Doney et al. 2004; Olsen et al. 1986), as nucleic acids can be directly extracted from environmental samples and studied. DNA extracts from microbial communities contain extremely rich genetic information including phylogenetic diversity and potential metabolic capability (Torsvik et al. 2002; Venter et al. 2004). In addition, RNA-based studies allow the measurement of microbial activities *in situ*, which in turn, provides clues to the linkage between community structure and functionality, as well as responses to environmental conditions (Corredor et al. 2004; Wawer and Muyzer 1995). Therefore, it is easy to recognize why the recovery of high-quality nucleic acids from the environmental samples is crucial to culture-independent analyses.

Ideally, a good nucleic acid extraction method should meet the following criteria: (i) the extraction yields should provide sufficient material for manipulation,

(ii) the extraction efficiency should be nonbiased and comparable for different samples, and (iii) the extracted DNA and RNA should be intact and free of inhibitors to subsequent steps in the analysis. Although a variety of nucleic acid extraction methods has been described and used for bacterioplankton samples (Paul and Pichard 1995), most have focused on nucleic acid quality rather than on yields or constancy of extraction efficiency. This relative lack of consideration of quantitative aspects of nucleic acid extraction is problematic to microbial ecology studies. On one hand, relatively inefficient extraction methods may result in low yields or even loss of RNAs from rare members in a microbial community, and therefore cause the under-representation or under-estimation of the abundance or roles of these microbes (Eichler et al. 2004). On the other hand, if not properly normalized, varying extraction efficiencies among samples may lead to biased or even erroneous data interpretation in quantitative and comparative studies.

Following the development of accurate gene quantification techniques such as real-time PCR, there has been a growing interest in quantifying the abundance and gene expression levels of specific microbial groups in the environment. Suzuki and coworkers (Suzuki et al. 2001b; Suzuki et al. 2000) developed a series of real-time PCR (TaqMan) assays to measure the abundance of 16S rRNA genes and performed a quantitative mapping of bacterioplankton population in Monterey Bay waters. More related to this thesis, are emerging gene expression studies based on mRNA quantification (Bustin 2002; Corredor et al. 2004; Wawrik et al. 2002; Wilson et al. 1999; Wyman 1999), which have been greatly facilitated and stimulated by the

discovery of millions of novel functional genes from the oceanic environment (Béjà et al. 2000a; Venter et al. 2004).

Gene and mRNA quantification in the environment can be carried out in a "relative" or "absolute" fashion. In most previous studies of microbes in the environment, target rRNA genes and target mRNA molecules were quantified relative to community-wide parameters such as total bacterial rRNA genes and total RNA amount, respectively. However, when using relative quantification methods, increases (or decreases) in target amounts might not be a direct reflection of actual target abundance change, but might instead be caused by the fluctuation of the abundance of non-target organisms. Absolute gene (or mRNA) quantification is a more effective alternative that can circumvent data misinterpretation due to changes in microbial community composition, where total gene (or mRNA) abundances are normalized to a sample volume or mass.

Uncertainty about nucleic acid extraction efficiency is a major factor precluding absolute quantification in environmental samples. In theory, nucleic acid extraction efficiency could be as high as 100%, target DNA or RNA molecules could be calculated as copies per ml seawater and these absolute values could be used to compare any two samples. In reality, it is highly unlikely that values of extraction efficiency will ever reach 100% and thus, constant extraction efficiency among samples can work an alternative to the full recovery of nucleic acids, allowing meaningful and a comparable estimate of gene copies per unit of seawater in a "quasi-absolute" fashion.

In this study, we developed a protocol for collection, storage and extraction of RNA from marine and estuarine bacterioplankton samples. RNA extraction is more challenging than DNA extraction. First of all, prokaryotic mRNA molecules are known for their relatively short half-life, which requires fast processing of samples and so, in order to decrease sample-processing time, we filtered relatively small volumes of water. Moreover, RNA molecules especially mRNA molecules are susceptible to ribonuclease (RNase) degradation during sample collection, storage and extraction, and thus, our samples were stored in a solution inhibitory to RNases. In order to optimize RNA yields, we compared five different commercially available beads used for cell disruption, since preliminary tests suggested that RNA might be lost through adsorption during bead beating (M.T. Suzuki, unpublished results). We also tested constancy of RNA extraction efficiency by estimating and comparing RNA (and mRNA) concentrations in the seawater samples, with the rationale that were the extraction efficiency constant among subsamples, the RNA (mRNA) concentrations estimated from different subsamples should be the same.

For the DNA extraction and quantification (see chapter 2), we followed previously developed procedures (Suzuki et al. 2001b), for which the constancy of DNA extraction efficiency has already been validated (Suzuki *et al.*, 2003 Aquatic Sciences Meeting, Salt Lake City, UT).

Materials and methods

Sample collection.

In order to compare total RNA yields when using five different types of beads for cell disruption, 15 replicated surface water bacterioplankton samples from the

Patuxent River Estuary (50 ml) were collected at the research pier of Chesapeake Biological Laboratory (38º19'N, 76º27.1'W), Solomons, MD, on October 10, 2003.

Constancy of RNA extraction efficiency among subsamples was initially tested by extracting bacterioplankton RNA from water samples from the Patuxent River Estuary collected at the research pier of Chesapeake Biological Laboratory, Solomons, MD. Eighteen bacterioplankton subsamples were collected on September 12, 2003, 18 subsamples on November 8, 2003, and another 24 subsamples on December 5, 2003, representing triplicates of 6 (10 ml, 50 ml, 100 ml, 200 ml, 400 ml, 500 ml), 6 (5 ml, 10 ml, 20 ml, 40 ml, 80 ml, 100 ml), and 8 (10 ml, 20 ml, 50 ml, 75 ml, 100 ml, 125 ml, 200 ml, 400 ml) different subsample sizes, respectively. Finally, additional triplicate 400 ml subsamples were collected on December 5, 2003 to evaluate possible RNeasy® spin column overloading.

A second test of constancy of RNA extraction efficiency was performed for oceanic water samples. Fifteen bacterioplankton subsamples were collected by filtering increasing amounts of surface water (500 ml, 1 l, 2 l, 4 l, 5 l, each in triplicate) collected by C.M. Preston on December 20, 2003 at the Hawaii Ocean Time Series (HOTS) station ALOHA (22º44´N, 158º2´W) aboard the R/V Kilo Moana.

Briefly, for all above-mentioned subsamples, water was pre-filtered through 1.6 µm GF/A filters (Whatman, Maidstone, UK) and filtered through 25 mm 0.2 µm polysulfone filters (Supor200®, Pall Gelman Inc., East Hills, NY). The Supor200® filters were then transferred to screw cap tubes containing 250 µl RNAlater (Ambion Inc., Austin, TX). Patuxent Estuary filters were frozen and stored at −70ºC

immediately after water filtration while the HOTS samples were frozen and kept at –20ºC aboard the R/V Kilo Moana before shipped frozen to Solomons where they were stored at –70ºC until RNA extraction procedures.

RNA extraction.

Total RNA was extracted from Supor200® 0.2 µm filters following a protocol adapted from the Qiagen RNeasy® 96 manual. Seeking the highest possible RNA yields, five different types of grinding beads were initially tested in triplicate RNA extractions from the 15 replicated filters and the total RNA extracted from each of these filters was measured. Bead types included Glass (100 µm, Glen Mills Inc., Clifton, NJ), Stainless Steel (100 µm, Glen Mills Inc.), Zirconia Silica (100~200 µm, Glen Mills Inc.), DNase/RNase treated Zirconium Oxide (200~400 µm, OPS Diagnostics, Bridgewater, NJ), and Low Protein Binding Zirconium Oxide beads (200 µm, OPS Diagnostics, named LPZ hereafter). Briefly, the filters in screw cap tubes were thawed and 320 µl of dry grinding beads (measured by filling a thin wall PCR tube) were added to each screw cap tube. ß-Mercaptoethanol was added in a 1:100 ratio to RLT buffer (Qiagen) and 875 µl of the mixture was added to each of the screw cap tubes. The tubes were then beaten in a MM301 mixer mill (RETSCH GmbH & Co., Haan, Germany) at maximum speed (30.0 Hz) for 2 min, followed by a 5-min incubation at 70° C. Exactly 800 µl of the liquid phase were transferred to new low-RNA-binding 2-ml microcentrifuge tubes (Ambion), avoiding transferring the beads. 800 µl of 100% ethanol were added to these tubes and well mixed with the supernatant, resulting in 1.6 ml of crude lysate in each tube, 800 µl of which were transferred to wells of a 96-well RNeasy® plate. The RNeasy microplate was then

sealed with AirPore tape to avoid cross contamination and centrifuged in a Sorvall®

Legend™ T Centrifuge with a Highplate® rotor (Kendro, Asheville, NC) at 5500 RCF

for 4 min at room temperature. The remaining crude lysate was added to the

corresponding wells of the 96-well RNeasy® plate and the centrifugation repeated.

The columns were washed with buffer RW1 and RPE buffer (Qiagen) at 5500 RCF

for 4 min at room temperature following the RNeasy® 96 kit protocol. RNA was

finally eluted from each column with 35 µl Diethylpyrocarbonate (DEPC) treated

water twice, resulting approximately in 60 µl RNA extracts. After the observation

that LPZ beads yielded the largest amount of total RNA, all subsequent RNA

extractions were performed as described above using LPZ beads.

Since RNA extractions from some September and November 2003

subsamples that were collected from larger amounts of water indicated a possible

overloading of RNeasy spin columns, three additional RNA extractions were

performed for triplicates filtered from 400 ml of December 5, 2003 subsamples,

where only the first 800 µl of the crude lysate was loaded to the spin columns, and the

yields of eluted RNA were compared to those from 400 ml subsamples, where the

total 1.6 ml crude lysate was loaded.

RNA quantification.

In order to remove DNA co-extracted with RNA molecules, RNA samples

collected from Station ALOHA were treated with Deoxyribonuclease (DNase) I using

the DNA-free™ kit (Ambion), according to the manufacturer's specification. This

further RNA purification allowed a more accurate RNA quantification, and is

required for the subsequent reverse transcription (RT) PCR procedure. RNA samples

from the Patuxent estuary were not DNase treated, and thus all concentrations and yields are slightly overestimated. RNA samples were fluorometrically quantified by RiboGreen® (Molecular Probes, Eugene, OR) staining on a Spectra MAX Gemini microplate spectrofluorometer (Molecular Devices, Sunnyvale, CA), following a protocol modified from the FluorImager™ nucleic acid quantification manual (Molecular Dynamics, Sunnyvale, CA). First, to generate a standard curve with the RNA standard (provided with the RiboGreen® kit), duplicate dilution sets of an RNA standard in DEPC-treated water were prepared with concentrations ranging from 10 000 to 52.5 pg/µl and a dilution factor of 0.35, and the extracted RNA samples were diluted 1:10 in DEPC treated water. RiboGreen® fluorescent dye was diluted 1:200 and 50 µl aliquots were distributed into wells of a low-fluorescence NUNC™ 96-well microtiter plate (Nalge Nunc International, Rochester, NY). Fifty microliters of RNA standards and diluted samples were added to the RiboGreen® dye, gently mixed and incubated in room temperature for 5 min before the fluorescence signals were measured using the PicoGreen® quantification software module (excitation wavelength=485 nm; emission wavelength=538 nm) shipped with the instrument.

Evaluation of constancy of RNA extraction efficiency.

Bacterioplankton samples from surface waters of the Patuxent River Estuary and station ALOHA were used to evaluate constancy of total RNA and PR mRNA extraction efficiency. Total RNA was extracted following the protocol described above using LPZ beads that were found to yield the highest amounts of RNA (see results) and RNA concentration was measured as described above. We performed a regression analysis to explore the relationships between sample size (filtered water

volume) and RNA concentration in extracts. We also calculated RNA concentration in the original water sample by dividing total RNA extracted by the volume of water filtered, allowing further comparison of extraction efficiencies among samples by ANOVA. Extraction efficiencies for the Patuxent estuary samples were calculated assuming 100% extraction efficiency for the subsamples yielding the highest average RNA concentrations. In addition, PR mRNAs from station ALOHA samples were reverse-transcribed to PR cDNAs and quantified using TaqMan assays (see chapter 3). The linear range for PR mRNA extraction was evaluated as for total RNA.

Results

RNA extraction efficiency.

Among the five different types of grinding beads tested for total RNA extraction, the LPZ beads showed the highest RNA yield extracting about 2.41 ± 0.18 µg total RNA (Table 1.1), followed by the Zirconia-Silica beads and Zirconium Oxide (DNase/RNase treated) beads, which gave about 65% of the highest yield. The remaining two, Glass beads and Stainless Steel beads, had much lower RNA yields, about 30% and 44% of that from LPZ beads, respectively.

Table 1.1. Summary of total RNA yields using 5 different types of grinding beads.

	Glass	Stainless Steel	Zirconia Sillica	DNase/RNase Zirconium Oxide	Low Protein Binding Zirconium Oxide
Average total RNA yields (µg)	0.73	1.05	1.59	1.57	2.41
Standard deviation (n=3)	0.36	0.21	0.29	0.05	0.18
Percentage of the highest (%)	30.2	43.7	66.0	65.1	100.0

It should be pointed out that the measured RNA yields were slightly overestimated since RNA extracts for this bead testing experiment were not DNase-treated. However, these values would likely be proportional to DNase-treated RNA yields and thus still comparable. Furthermore, measurements of total RNA concentrations of station ALOHA samples before and after DNA-*free*™ treatment showed that although RNA concentrations were overestimated due to the co-extraction of DNA, DNase-treated RNA amounts were proportional to untreated ones (data not shown).

Constancy of total RNA extraction efficiency.

Preliminary results from Patuxent Estuary bacterioplankton samples revealed that bacterioplankton samples yielding 300 to 2500 ng total RNA were extracted with a relatively constant efficiency, while bacterioplankton samples yielding less or more starting RNA were extracted with lower extraction efficiencies (Fig. 1.1D). RNA concentration of the September 2003 water sample estimated from different subsamples was not significantly different for 10 ml (yield 270 ± 65 ng RNA), 50 ml (yield 1600 ± 146 ng RNA) and 100 ml (yield 2450 ± 236 ng RNA) water (Fig. 1.1A,

ANOVA, p=0.25). Similarly, estimated RNA concentrations in the November 2003 water sample were not significantly different for 80 ml (yield 900 ±39 ng RNA) and 100 ml (yield 1200 ± 57 ng RNA) samples (Fig. 1.1B, ANOVA, p=0.08), and estimated RNA concentrations of the December 2003 water sample were not significantly different in 100 ml (yield 960 ± 286 ng RNA), 125 ml (yield 1000 ± 32 ng RNA) and 200 ml (yield 2100 ± 71 ng RNA) samples (Fig. 1.1C, ANOVA, p=0.26).
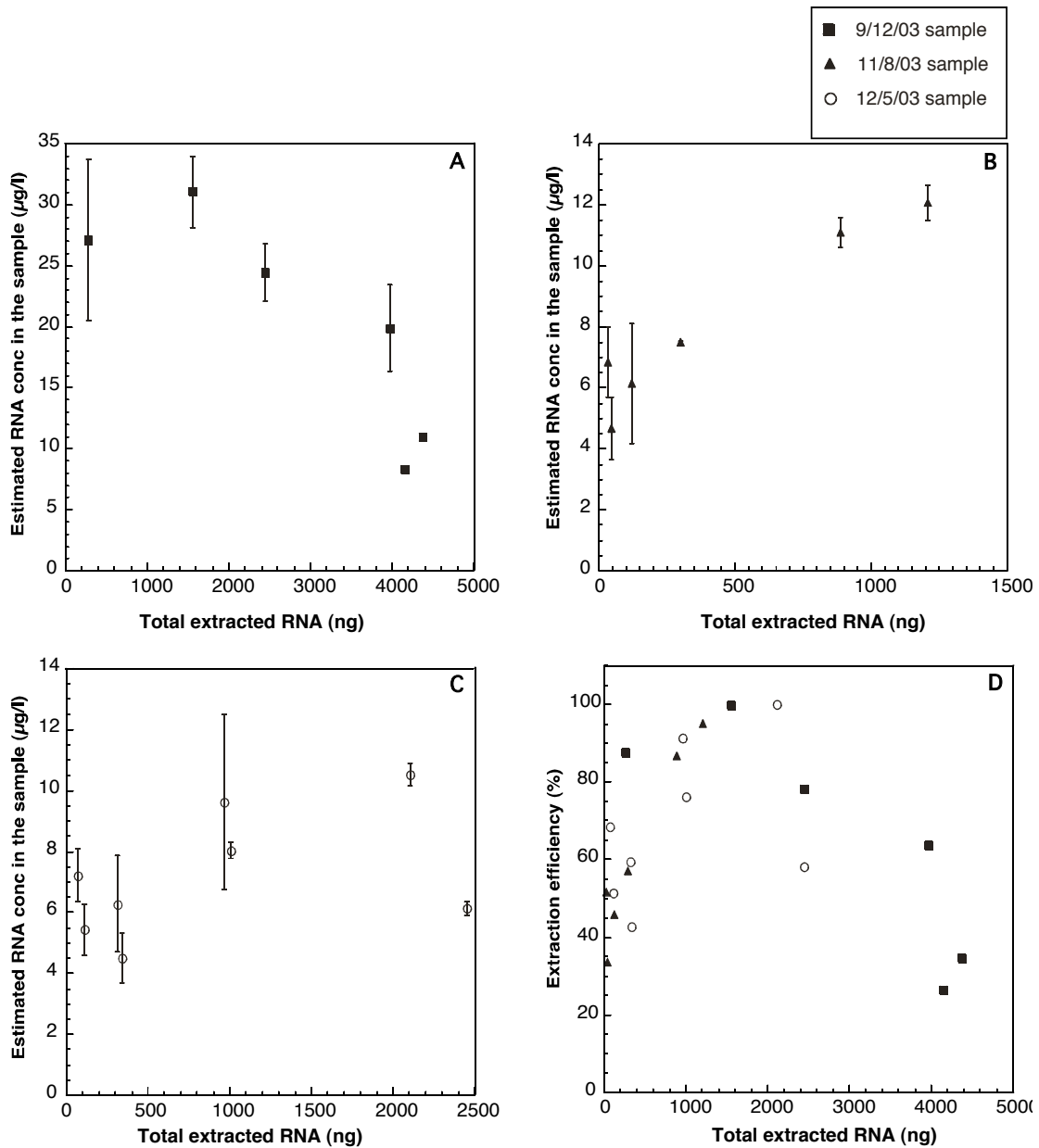
Fig. 1.1. Test of constancy of total RNA extraction efficiency for Patuxent Estuary samples. **(A)** RNA concentration in the September water sample estimated from six subsamples filtered from different water volumes. **(B)** RNA concentration in the November water sample estimated from six subsamples filtered from different volumes. **(C)** RNA concentration in the December water sample estimated from eight subsamples filtered from different volumes. **(D)** RNA extraction efficiencies calculated as a ratio of RNA yields to that of the sample with highest yield at each of the collection dates. All subsamples were triplicates (n=3) and error bars indicate one standard deviation.

21

Overloading of RNeasy columns above the 2500 ng RNA range was corroborated by a comparison of yields of RNA eluted from columns loaded with different volumes of crude lysate, where all other conditions were kept constant. An amount of 2000 ± 168 ng of RNA was eluted from columns loaded with 800 μl of crude lysate, while only 2500 ± 83 ng of RNA was eluted from columns loaded with total 1.6 ml of crude lysate, strongly indicating that RNA was not fully recovered from RNeasy® columns when more than 2500 ng of RNA was loaded.

For RNA extracted from Hawaii Ocean samples, we found a strong linear relationship between total extracted RNA (ng) and filtered water volume (l) with $R^2 =$ 0.93 (Fig. 1.2A). Figure 1.2B shows the estimated seawater RNA concentrations. Except for the concentration estimated from the 0.5 l water subsamples that were higher, no significant differences were found among 1, 2, 4, and 5 l samples (ANOVA, p=0.34). This indicated that extraction efficiency for total RNA at station ALOHA was not significantly different for water volumes between 1 and 5 l, which yielded 200 ± 70 ng to 1000 ± 145 ng of total extracted RNA.
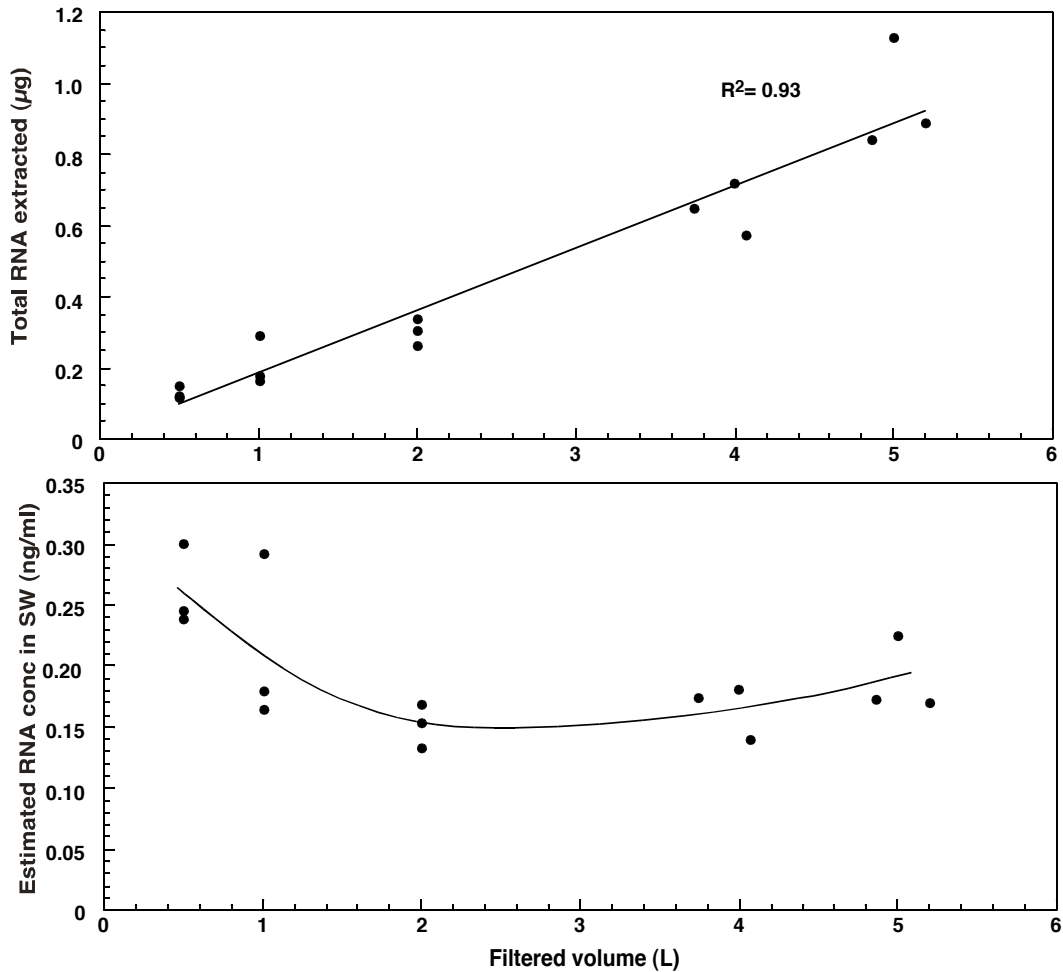
Fig. 1.2. Test of constancy of total RNA extraction efficiency for Station ALOHA samples. **(A)** total RNA extracted from subsamples filtered from increasing water volumes. The line represents a linear regression line. **(B)** RNA concentration in the seawater estimated from each of the subsamples. The curve is a fitting line through average values of the triplicates for each subsample.

Constancy of Proteorhodopsin mRNA extraction efficiency.

The R-square for the regression between total PR mRNA extracted and the filtered water volume was high (0.94) (Fig. 1.3A). However, the estimated PR mRNA concentration in the seawater increased as a function of increasing water volume before leveling off. PR mRNA concentrations estimated from the 4 l subsamples and 5 l subsamples were not significantly different (Fig. 1.3B, student's t,

23

t=2.78, p=0.37), indicating that the PR mRNA extraction efficiency was constant in samples yielding 650 ± 75 ng to 1000 ± 145 ng of total extracted RNA.
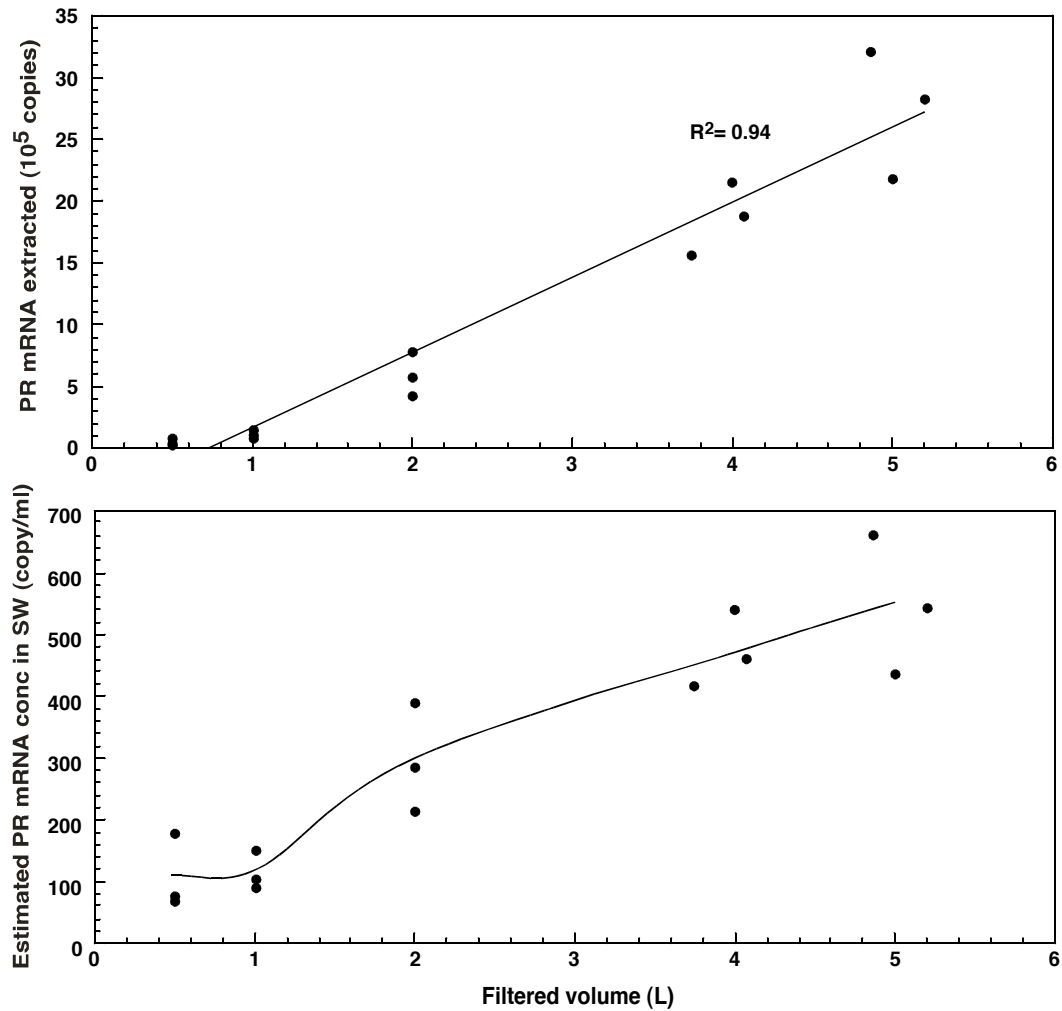
Fig. 1.3. Test of constancy of PR mRNA extraction efficiency for Station ALOHA samples. **(A)** PR mRNA (estimated from cDNAs assuming 100% reverse transcription efficiency) in RNA extracted from subsamples filtered from increasing volumes. The line represents the regression line. **(B)** PR mRNA concentration in the seawater (estimated from cDNAs assuming 100% reverse transcription efficiency) calculated from each of the subsamples. The curve is a fitting line through average values of the triplicates for each subsample.

Discussion

Extraction of high-quality nucleic acids is a key step to the successful application of molecular techniques in culture-independent microbial studies. In addition, for quantitative assays, the constancy of extraction efficiency is important since it allows comparison among samples through absolute (or "quasi-absolute") quantification. However, common RNA extraction methods involve organic phase separations (Phenol or Trizol) that in most cases do not retrieve standardized amounts of aqueous phase, introducing sample-to-sample variation in yields and extraction efficiencies. Moreover, RNA extraction yields have not been extensively investigated: few studies reported comparatively high extraction efficiency (64-87%) from bacterial isolates using RNA-specific extraction methods (Pichard and Paul 1991; Ward et al. 1995), and Pichard and Paul (Pichard and Paul 1993) showed that the change of filters from Whatman GF/F filters to Millipore Durapore filters greatly improved RNA extraction efficiency from *E. coli* cells, from 11% to 70% (Pichard and Paul 1993). The well known high binding of RNA to glass under chaotropic conditions might explain some of these results.

Filter type effects on RNA extraction efficiency were not tested in this study, but our RNA extraction experiments suggested that extraction efficiency could be improved (up to 3-fold) with a proper choice of bead type used for cell disruption. Extraction methods using bead beating to homogenize cells and liberate nucleic acids have the inherent problem that biochemical molecules can be adsorbed to bead surfaces, reducing the yields and affecting the extraction efficiencies. This problem is especially significant for small-sized samples like those used in the majority of

environmental microbial studies, since a larger proportion of extractable RNA would be lost by adsorption. Among the grinding beads we tested, the low protein binding beads gave the highest RNA yields, most likely because that the proprietary coating process minimized nonspecific binding of molecules. In contrast, glass beads, commonly used in RNA extraction protocols had a significantly lower yield, suggesting a higher adsorption of RNA molecules, and thus should be avoided in quantitative studies since this adsorption is likely to be concentration dependent.

Although the nominal binding capacity of RNeasy® spin columns is 100 μg of total RNA (QIAGEN RNeasy® handbook), our results suggested that the upper limit for linear elution of loaded RNA is much lower (less than 2.5 μg). The nominal binding capacity was determined using partially purified RNA molecules (according to Qiagen technical services) which, in contrast to RNAs released from cells without purification, contain little cell debris to interfere with RNA binding and elution. For non-purified RNA samples loaded to RNeasy® spin columns, the elution efficiency of the bound RNA molecules tends to vary with the changes of sample type (seawater or isolates, cell type, etc.), separation technology used (vacuum or spin, etc.), or experimental settings (temperature, spin speed, vacuum force, elution volume, etc.). Therefore, in order to avoid underestimation of total RNA yields, possible RNeasy® column overloading should be tested for RNeasy®-based RNA extractions, especially in quantitative and comparative assays.

Accurate quantification of total RNA is important to the assessment of RNA extraction efficiency, particularly in gene expression studies that normalize target mRNA to total RNA (Bustin 2002; Leclerc et al. 2002). RiboGreen assays, which

have a linear dynamic range from 1 ng/ml to 1 µg/ml RNA, are suggested to be 200-fold more sensitive than ethidium bromide assays, and 1000-fold more sensitive than UV absorbance measurements. However, since RiboGreen also binds to DNA and the amount genomic DNA contaminants in RNA extracts could reach the same magnitude as that of certain mRNA (Vandecasteele et al. 2001), the removal of contaminated genomic DNA molecules from RNA extracts is essential.

We showed that within a certain range of total extracted RNA (0.3 to 2.5 µg for the Patuxent Estuary samples and above 0.2 µg for Station ALOHA samples), extraction efficiency was relatively constant among subsamples filtered from different sample volumes (Fig. 1.1 and Fig. 1.2). Bostrom and coworkers (Bostrom et al. 2004) have previously suggested that sample volume affects DNA extraction efficiency, and here we also show that water volume filtered has an effect on RNA extraction efficiency. The significantly higher total RNA extraction efficiency observed for Station ALOHA subsamples filtered from the smallest water volumes (0.5 l) was somewhat unexpected since we observed lower extraction efficiencies for smaller subsample in Patuxent Estuary samples. Furthermore, as mentioned above, because a proportionally larger fraction of total RNA probably would be adsorbed or lost during extraction, lower RNA extraction efficiency would be predicted for smaller samples size. Possible explanations for this observation are the combined effects of 1) a more effective lysis of cells at lower numbers, 2) lesser amount of substances affecting extraction for samples from the open ocean, and 3) better elution of total RNA from the silica columns for smaller total RNA samples. Furthermore, the discrepancy in the ranges where total RNA extraction efficiency was constant

between Patuxent Estuary samples and Station ALOHA samples were very likely due to the lack of DNase-treatment of the former.

An important observation was the discrepancy between the ranges where total RNA and PR mRNA extraction efficiencies were constant. This discrepancy of extraction could be explained by the lower extraction efficiencies of abundance of target mRNAs compared with rRNAs and tRNAs, although we were not able to find evidence for this effect in the literature. More likely, lower reverse transcription rates for mRNAs at lower concentrations, also suggested by previous studies (Curry et al. 2002) could be causing this apparent lower efficiency of extraction.

We did not test the integrity of RNA molecules extracted in this study due to the following considerations. First, the Qiagen RNeasy protocol has been widely used and recognized as to yield high quality RNA products (in terms of purity and integrity). Second, oceanic waters are highly oligotrophic and the amount of RNA extracted from large volumes of water samples is small, compared to the relatively large amount of RNA (1~3 μg, or about the entire sample) needed to run formaldehyde agarose gels to test for RNA integrity. Finally, thanks to the specific design of TaqMan assays used in this study (see Chapter 3), we were able to reverse-transcribe only a small region of mRNA sequences instead of the complete mRNA molecules and quantify relatively small fragments (109-bp in this study) of the target cDNA molecules. Therefore, partial mRNA degradation would be somewhat acceptable, since they would still be detected and quantified as long as the regions between the primers are intact.

In conclusion, we developed an RNA collection, storage and extraction strategy for marine and estuarine bacterioplankton yielding RNA in concentrations and quality suitable for reverse transcription and relatively constant extraction efficiencies allowing inter-sample comparisons. Combined with a previously optimized DNA extraction method (Suzuki *et al.*, 2002 Aquatic Sciences Meeting, Salt Lake City, UT), this method allows "quasi-absolute" quantitative assays for gene (and mRNA) abundance, as well as normalized gene expression measurements (Chapter 3). Normalized measurements of gene expression in turn provide a linkage between microbial community structure and *in situ* activities in the content of environmental conditions. Most importantly, this optimized nucleic acid extraction method provides researchers with a possibility to study activities of not yet cultured microbes in the aquatic ecosystems.

Literature cited

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN and DeLong EF (2000a). "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea." Science 289(5486): 1902-6.

Bostrom KH, Simu K, Hagstrom A and Riemann L (2004). "Optimization of DNA extraction for quantitative marine bacterioplankton community analysis." Limnology and Oceanography-Methods 2: 365-373.

Bustin SA (2002). "Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems." J Mol Endocrinol 29(1): 23-39.

Corredor JE, Wawrik B, Paul JH, Tran H, Kerkhof L, Lopez JM, Dieppa A and Cardenas O (2004). "Geochemical rate-RNA integrated study: Ribulose-1,5-bisphosphate carboxylase/oxygenase gene transcription and photosynthetic capacity of planktonic photoautotrophs." Appl Environ Microbiol 70(9): 5459-5468.

Curry J, McHale C and Smith MT (2002). "Low Efficiency of the Moloney Murine Leukemia Virus Reverse Transcriptase during Reverse Transcription of Rare t(8;21) Fusion Gene Transcripts." Biotechniques 32(4): 768-775.

Doney SC, Abbott MR, Cullen JJ, Karl DM and Rothstein L (2004). "From genes to ecosystems: the ocean's new frontier." Front Ecol Environ 2(9): 457-466.

Eichler S, Weinbauer MG, Dominik K and Hofle MG (2004). Extraction of total RNA and DNA from bacterioplankton. Molecular Microbial Ecology Manual. van Elsas, JD. Dordrecht, the Netherlands, Kluwer Acad. Publ.: 1780.

Leclerc GJ, Leclerc GM and Barredo JC (2002). "Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines." Cancer Cell Int 2(1): 1.

Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR and Stahl DA (1986). "Microbial ecology and evolution: a ribosomal RNA approach." Annu Rev Microbiol 40: 337-65.

Paul JH and Pichard SL (1995). Extraction of DNA and RNA from Aquatic Environments. Nucleic Acids in the Environment. van Elsas, JD. New York, Springer: 153-177.

Pichard SL and Paul JH (1991). "Detection of Gene-Expression in Genetically Engineered Microorganisms and Natural Phytoplankton Populations in the Marine-Environment by Messenger-Rna Analysis." Appl Environ Microbiol 57(6): 1721-1727.

Pichard SL and Paul JH (1993). "Gene-Expression Per Gene Dose, a Specific Measure of Gene-Expression in Aquatic Microorganisms." Appl Environ Microbiol 59(2): 451-457.

Suzuki MT, Taylor LT and DeLong EF (2000). "Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays." Appl Environ Microbiol 66(11): 4605-14.

Suzuki MT, Preston CM, Chavez FP and DeLong EF (2001b). "Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay." Aquatic Microbial Ecology 24(2): 117-127.

Torsvik V, Ovreas L and Thingstad TF (2002). "Prokaryotic diversity--magnitude, dynamics, and controlling factors." Science 296(5570): 1064-6.

Vandecasteele SJ, Peetermans WE, Merckx R and Van Eldere J (2001). "Quantification of expression of Staphylococcus epidermidis housekeeping genes with TaqMan quantitative PCR during in vitro growth and under different conditions." <u>J Bacteriol</u> 183(24): 7094-101.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH and Smith HO (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." <u>Science</u> 304(5667): 66-74.

Ward DM, Ruff-Roberts AL and Weller R (1995). Methods for extracting RNA or ribosomes from microbial mats and cultivated microorganisms. <u>Molecular Microbial Ecology Manual</u>. van Elsas, JD. Dordrecht, the Netherlands, Kluwer Acad. Publ.: 1780.

Wawer C and Muyzer G (1995). "Genetic diversity of Desulfovibrio spp. in environmental samples analyzed by denaturing gradient gel electrophoresis of [NiFe] hydrogenase gene fragments." <u>Appl Environ Microbiol</u> 61(6): 2203-10.

Wawrik B, Paul JH and Tabita FR (2002). "Real-time PCR quantification of rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes." <u>Appl Environ Microbiol</u> 68(8): 3771-9.

Wilson MS, Bakermans C and Madsen EL (1999). "In situ, real-time catabolic gene expression: Extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater." <u>Appl Environ Microbiol</u> 65(1): 80-87.

Wyman M (1999). "Diel rhythms in ribulose-1,5-bisphosphate carboxylase/oxygenase and glutamine synthetase gene expression in a natural population of marine picoplanktonic cyanobacteria (Synechococcus spp.)." <u>Appl Environ Microbiol</u> 65(8): 3651-9.

CHAPTER 2: Phylogenetic Analysis of Actively Expressed Proteorhodopsins at Station ALOHA

Abstract

A cDNA clone library was constructed using primers targeting proteorhodopsin (PR) from a RNA sample colleted from 25 m depth at Station ALOHA in the North Pacific Gyre. Ninety-six clones were sequenced and 62 were identified to contain proteorhodopsin-coding sequences. Forty-one unique PR amino acid sequences were combined with sequences deposited to GenBank and novel sequences retrieved from the Mediterranean and Red Seas, and a phylogenetic analysis of these sequences suggested that the PR cDNA sequences originated from diverse and distinct groups. We were able to assign 33 of the 41 unique sequences to *Alphaproteobacteria* and 1 to *Gammaproteobacteria*, with the remaining 7 sequences unassigned but related to unidentified PR sequences retrieved from a Sargasso Sea shotgun library. Interestingly, green-light absorbing PRs were absent in this clone library, even though they were the only PR types retrieved in a cDNA library constructed using a previously published and less degenerate primer set targeting PR genes. In contrast, the most prevalent group in this PR cDNA library were closely related to the PR harbored by the *Pelagibacter ubique* (SAR11) group, suggesting that SAR11 or closely-related bacteria may represent a significant fraction of bacteria that harbor and actively express PR genes in the photic zone of North Pacific Central Gyre.

Introduction

Proteorhodopsins (PRs), as homologs of archaeal bacteriorhodopsins (BR), are bacterial retinal-binding proteins that function as proton pumps to produce cellular energy from light in marine ecosystems. The first PR-encoding gene was

discovered in uncultured *Gammaproteobacterium* (later assigned to SAR86-II clade, Suzuki et al. 2001a) in Monterey Bay surface waters, California (Béjà et al. 2000a). Recent cultivation-independent molecular surveys have revealed that different SAR86 subgroups harbor divergent PR genes (Sabehi et al. 2004; Sabehi et al. 2003), and that planktonic *Alphaproteobacteria* also contain PR genes (de la Torre et al. 2003). Furthermore, a much higher diversity of PR genes was indicated by the discovery and analysis of over 600 novel PR genes from a Sargasso Sea shotgun library (Venter et al. 2004). In addition to the high diversity and abundance of PR-containing bacteria has also been suggested (ca. 13% of the microorganisms in the photic zone, according to a recent paper by Sabehi and coworkers, Sabehi et al. 2005). However, these previous studies were mainly focused on the presence of PR genes while little is known about the abundance and phylogenetic affiliations of actively expressed PR genes in the natural ecosystems.

Reverse transcription (RT) of RNA to complementary DNA (cDNA), followed by polymerase chain reaction (PCR) amplification of cDNA is considered one of the most sensitive method for detecting mRNA and monitoring actively expressed genes (Fey et al. 2004; Sharkey et al. 2004a; Wawrik et al. 2002; Wilson et al. 1999). However, the production of cDNA in the RT stage is still not thoroughly understood and believed to be a major variation source in gene expression analyses (Ståhlberg et al. 2004b). This is particularly critical for environmental RNA samples, in which specific mRNAs are in relative low abundance and high heterogeneity.

Three factors have been shown to mainly affect the specificity, reproducibility and efficiency of a RT reaction: the structure and abundance of mRNA templates, the

primers used for cDNA synthesis, and the properties of reverse transcriptases (RTases). Intramolecular folding and heteroduplex formation of mRNA molecules may block access of primers to the target sites and is a template-dependent and inherent problem (Sohail and Southern 2000). Also, low abundance of mRNA molecules tend to limit mRNA-to-cDNA conversion efficiency, reducing the sensitivity of RT-PCR assays (Curry et al. 2002). Fortunately, this problem can be alleviated by the appropriate choice of primers and RTases. Ståhlberg and coworkers have presented a detailed comparison of cDNA yields between RT reactions using four types of primers (Ståhlberg et al. 2004a) and eight types of reverse transcriptases (Ståhlberg et al. 2004b). In general, the use of target-specific primer and thermo-stable enzymes allows priming of specific mRNA at elevated temperatures with decreasing secondary structure and increasing priming specificity and efficiency of RT reactions. Nevertheless, since optimal conditions of reverse transcription are highly target-dependent (Ståhlberg et al. 2004b), no single RT reaction condition is optimal for all target genes.

To obtain an overall picture of cDNA sequence diversity in mixed populations, a cDNA clone library is often constructed. A custom cDNA library for a single target gene contains, in theory, cDNAs representing mRNAs present *in situ* at the sampling time and allows further analyses by sequencing and phylogenetic comparisons. Compared with genomic libraries, cDNA libraries provide insights into the diversity and quantity of actively expressed genes in addition to physically present genes. For example, Béjà and coworkers (Béjà et al. 2002) employed RT-PCR to construct a *pufM* (genes coding the M subunit of the Photosynthetic Reaction Center

of *Bacteria*) cDNA library and identified groups actively expressing photosynthetic genes in natural microbial populations in Monterey Bay, California.

In the current study, we reverse-transcribed PR mRNA from a RNA sample from 25 m depth water collected at station ALOHA in the North Pacific Gyre using degenerate primers, and constructed a PR cDNA library from amplified PR cDNAs. Ninety-six clones were sequenced in an attempt to identify the actively expressed PR genes in the North Pacific Gyre, and to investigate their phylogenetic affiliations to previously sequenced PRs. These results were also used to guide our target choice for the study of *in situ* diel dynamics of PR expression, discussed in Chapter 3.

Materials and methods

Sample collection.

Seawater was collected from 25 m depth at the Hawaii Ocean Time Series station ALOHA (22º44´N, 158º2´W) on 13 December 2002.

Bacterioplankton from about 3800 ml of seawater was collected onto a 25 mm 0.2 μm polysulfone filter (Pall Gelman Inc.), frozen and kept at –20ºC aboard the R/V Kilo Moana before shipped frozen to Solomons where it was stored at –70ºC until RNA and DNA extraction procedures. Total RNA extraction, DNA removal and RNA quantification were performed following the procedures described in Chapter 1.

Reverse transcription and cDNA amplification.

M-MLV Reverse Transcriptase (Invitrogen, Carlsbad, CA) and ThermoScript™ RNase H⁻ Reverse Transcriptase (Invitrogen) were evaluated for the reverse transcription of PR mRNA from total RNA samples. In addition, we compared reverse transcription reactions using random hexamers (Promega, Madison,

WI) and two target-specific primers. The two target specific primers were the reverse primers in PCR primer sets designed by G. Sabehi and O. Béjà. These primer sets were: Primer Set A: Prr9F (5'-ACCATGGGTAAANNATTANTGATNTTAG-3') and PRR (5'-AGCATTAGAAGATTCTTTAACGC-3') (Sabehi et al. 2003), and Primer Set B: RYIDWLfwd (5'-MGNTAYATHGAYTGGYT-3') and GWAIYPrv (5'-GGRTADATNGCCCANCC-3') (Sabehi et al. 2005). Primer Set B was more degenerate than the Primer Set A. All reverse transcription reactions were performed in a total reaction volume of 10 µl according to the corresponding RT protocols (for the ThermoScript™ RTase, samples were incubated at 52ºC for reverse transcription). The PR cDNA product was subsequently amplified by PCR with the corresponding degenerate primer set. In a final volume of 25 µl, PCR mixtures contained 2.5 µl of PCR 10X buffer (Invitrogen, Carlsbad, CA), 0.2 mM of dNTPs (Promega, Madison, WI), 3 mM of MgCl2 (Invitrogen), 0.5 µM of primers, 2.5 µl of PR cDNA product and 0.05 U µl^{-1} of Platinum® Taq DNA polymerase (Invitrogen). Thermal cycle parameters were as follows: 4 min at 92°C and 40 cycles of 1 min at 92°C, 1 min at 52°C, and 1 min at 72°C. PCR products were loaded in a 1% agarose gel containing 0.5 µg/ml Ethidium Bromide (EB), and electrophoresed at 70 V for 1.5 h. Since the amplified DNA was later released from gel for the cloning procedure, modified Tris-acetate EDTA (TAE) buffer (1 mM NaEDTA (pH 8.0), 40 mM Tris-acetate) was used to prepare the agarose gel and as the running buffer. The DNA band of the targeted size was carefully cut from the gel using a Dark Reader Transilluminator (Clare Chemical Research, Dolores, CO) and DNA isolated from gel using an *Ultrafree*-DA centrifuge filter (Millipore, Billerica, MA).

Construction and sequencing of PR cDNA library.

Purified DNA was ligated with the pCR®4-TOPO® vector following the TOPO TA Cloning® kit protocol (Invitrogen). Ionic strength of ligation products was decreased by drop dialysis as previously described (Béjà et al. 2000b) and products were used to transform into *E.coli* DH10B™ electrocompetent cells by electroporation in 1 mm gap cuvettes using a Gene Pulser (Bio-Rad, Hercules, CA) set at 2500 V, 25 µF and 100 ohms. After Electroporation, *E.coli* cells were recovered overnight at 37º C at 225 rpm in a shaker incubator (Lab-Line® Orbit-Environ Shaker, Lab-Line, Melrose Park, IL), and 100 µl were plated onto LB agar plates with 50 µg/ml of Kanamycin. After overnight incubation at 37º C, 96 colonies were picked from the plates and transferred to 96 wells of a microtiter plate in 180 µl LB media with 7% glycerol and 50 µg/ml of Kanamycin. The plate was incubated overnight at 37º C and stored at –70º C until further manipulation.

Plasmids from all 96 clones from the PR cDNA clone library were purified using the SprintPrep™ plasmid purification kit (Agencourt, Beverly, MA) following the manufacturer's instructions, and 1 µl of the purified plasmids was used as template in cycle sequencing reactions. A 10 µl reaction contained 1 µl of template, 0.5 µl of BigDye® terminator v3.1, 1.75 µl of 5X BigDye® buffer, 1 µl of M13F primer (3.2 µM) and 5.75 µl of DEPC-treated water. 40 cycles of 10 sec at 96º C, 5 sec at 50º C, and 2:30 min at 60º C were run on an AB9700 thermal cycler (Applied Biosystems, Foster City, CA). The cycling sequencing products were purified with the CleanSeq™ dye-terminator removal kit (Agencourt) and subsequently sequenced on an AB3100 Genetic Analyzer (Applied Biosystems).

Phylogenetic analysis of PR genes.

PR gene sequences previously deposited in the GenBank (as of July, 2003) including those from the Sargasso Sea shotgun clone library (Venter et al. 2004) were retrieved and combined with the PR cDNA sequences obtained in this study. Nucleotide sequences were imported into a sequence database using the ARB software (Ludwig et al. 2004) and translated into amino acids. The amino acid sequences were preliminarily aligned using the ClustalW alignment tool in ARB. Using this alignment, a Neighbor-Joining (NJ) guide tree was generated in ARB, and 93 full-length sequences representative of most sequence types in the guide tree were exported, aligned using T_Coffee (Notredame et al. 2000) and re-imported into the ARB database. Each of the 93 sequences aligned by T_Coffee was used to re-align close related sequences based on the guide tree using the ARB alignment tool, followed by a manual clean up of the resulting alignment. Thirty-two unique novel PR DNA sequences obtained from BAC libraries constructed from Red and Mediterranean Sea bacterioplankton samples (provided by G. Sabehi and O. Béjà) and the sequence of a PR protein "protein A" (Stapels et al. 2004) recently identified from Marine Bacterium A (a cultured *Pelagibacter ubique* strain; S.J. Giovannoni, Personal communication) were added to the alignment based on the alignment of close related sequences (as determined by a nucleotide based PT_SERVER), using the ARB alignment tool. Based on this final PR protein alignment, 89 amino acid alignment positions (position 150 to 193 and position 211 to 255 in the amino acid sequence of EBAC31A08 (Béjà et al. 2000a, accession Number: AF279106) from 272 sequences were selected for all phylogenetic reconstructions. These positions

were selected based on the criteria that these positions were present in the vast majority of unique sequences and showed significant homology.

Distance-based and Bayesian methods were used to generate phylogenetic trees for PR sequences. For distance-based phylogenetic inference, Phylip package (J. Felsenstein, 1993, University of Washington, Seattle) implemented in ARB was employed to generate a NJ tree using the JTT model of amino acid substitution rates (Jones et al. 1992). Bootstrap analysis was performed with 100 randomly resampled datasets to estimate the robustness of the tree topology. In addition, we used MrBayes 3.0 (Huelsenbeck et al. 2001; Ronquist and Huelsenbeck 2003) to generate a Bayesian tree using the WAG amino acid substitution rate model (Whelan and Goldman 2001) with among-site rate heterogeneity corrected using the discrete gamma distribution model (category k=4) (Yang 1994, 1996). Four parallel chains of 500,000 generations were run, trees were sampled every 100 generations, and a "burnin" of 1500 trees was excluded when generating the consensus tree.

Results

Assessment of RT primers and enzymes.

Our results indicated that the choice of primers and enzymes greatly influenced the efficiency as well as specificity of RT reactions. Table 2.1 lists the results of RT-PCR reactions using different combination of primers and RTases, and subsequent sequencing of PR cDNA clones. We were only able to detect PCR products with targeted sizes in agarose gels when specific primers were used in the RT step, while no bands were detected when random hexamers were used for cDNA synthesis. RT reactions catalyzed by the M-MLV RTase resulted on the

amplification and cloning of non-specific cDNAs transcribed from rRNA instead of proteorhodopsin. In contrast, reactions using the thermo-stable ThermoScript™ RTase (at a higher temperature) greatly improved the specificity of RT reactions, leading to the amplification and cloning of PR cDNAs.

After the sequencing of 16 clones and the subsequent screening of 96 clones using Restriction Fragment Length Polymorphism (RFLP) analysis (data not shown), it was evident that the PR cDNA clone library constructed using primer set A were very closely related and belonged to the Green-absorbing PR group (Béjà et al. 2001; Sabehi et al. 2003). Thus, in order to retrieve more diverse PR mRNA sequences, degenerate primer set B designed by Sabehi and Béjà (based on sequences obtained from the Sargasso Sea shotgun library (Venter et al. 2004) and all PR sequences deposited prior to the design of this primers) was used in RT PCR of PR mRNA, resulting in a much higher diversity of PR cDNA sequences (Fig. 2.1).

Table 2.1. Comparison of the efficiency and specificity of RT reactions using different primers and reverse transcriptases.

Primer and reverse transcriptase [a]	DNA band	Clone sequence
RP+MMLV	No	N/A
SP+MMLV	Yes	non-specific rRNA
RP+ThermoScript	No	N/A
SP+ThermoScript	Yes	PR (more diverse sequences when more degenerate SP was used)

[a] RP and SP represent random primer (hexamer) and specific primer, respectively.

Phylogenetic diversity of PR cDNAs from Station ALOHA.

Highly diverse PR cDNA sequences were retrieved in this study, as shown in the PR phylogenetic trees inferred using distance-based (Neighbor-Joining, Fig. 2.1A) and Bayesian (Fig. 2.1B) methods. PR cDNA sequences originated from divergent prokaryotes closely related to *Alphaproteobacteria* (e.g., HOT2C01, de la Torre et al. 2003) and *Gammaproteobacteria* (e.g., HOT4E07, Sabehi et al. 2004), as well as species whose phylogenetic identity is unknown (e.g. many of the Sargasso Sea shotgun sequences, Venter et al. 2004). Despite this high diversity, the predominant PR cDNA sequences from the cDNA clone library were found to be closely related to a PR ("protein A", Stapels et al. 2004) identified via proteome analysis of one cultured *Pelagibacter unique* (SAR11) strain (S.J. Giovannoni, personal communication). Among 62 clones containing PR cDNAs, 40 clones (with 27 unique amino acid sequences) belonged to Clade cDNA1 containing *P. ubique* and supported by a bootstrap value of 85 (NJ, Fig. 2.1A) and confidence value of 77% (MrBayes, Fig. 2.1B). More notably, amino acid sequences of 7 of these clones were identical in a span of 110 amino acids (although they were not identical at the DNA level) and formed a PR subclade with 3 other close relatives. This PR subclade was chosen as the target for real-time PCR assays (Chapter 3).

As shown in Figure 2.1, the topology of the Bayesian tree was fairly similar to that of the NJ tree, with the following noteworthy exceptions: 1) In the Bayesian tree, the subclade targeted by the real-time PCR assays did not form a clade including SAR11 PR sequence, though these sequences were still closely related (>80% amino acid identity), 2) Group "Sargasso4" contained 5 PR shotgun sequences that formed a

separate clade in the NJ tree, and 3) 16 sequences including 2 cDNA sequences and 14 shotgun sequences representing a single clade in the NJ tree formed 4 different clades in the Bayesian tree. Besides, many additional minor differences in the placement of some PR sequences that did not significantly affect the general the tree topology were observed between the two trees.

Fig. 2.1. Phylogenetic reconstruction of PR proteins based on protein sequences. The SAR11 PR sequence identified by proteome analysis of a cultured *Alphaproteobacteria* strain is indicated with a star symbol. PR cDNA sequences and groups containing PR cDNA sequences obtained in this study are marked in boldface. The number in the parentheses after the PR cDNA-containing groups indicates the number of PR cDNA sequences contained in these groups. The scale bar represents the number of substitutions per sequence position. The subtree in the shaded area represents all sequences contained in the "cDNA1" group in the main tree. **(A)** Neighbor-Joining phylogenetic tree reconstructed using ARB. Bootstrap values above 40 over 100 trees are shown. **(B)** Bayesian phylogenetic tree reconstructed using MrBayes v3.0. Numbers on the nodes represent confidence values.

shotgun clone IBEA_CTG_SVAAR31TR
cDNA clone PRN04
cDNA clone PRNb53
cDNA clone PRNb55
cDNA clone PRNb74
cDNA clone PRNb52
shotgun clone IBEA_CTG_SZART11TR
shotgun clone IBEA_CTG_UEAG457TF
cDNA clone PRNb2
cDNA clone PRNb57
cDNA clone PRNb23
cDNA clone PRNb24
shotgun clone IBEA_CTG_UDAXV49TR
shotgun clone IBEA_CTG_1995897
shotgun clone IBEA_CTG_SSBQ730TF
shotgun clone IBEA_CTG_2125810
shotgun clone IBEA_CTG_2154055
shotgun clone IBEA_CTG_UDAFD69TR
shotgun clone IBEA_CTG_2085296
shotgun clone IBEA_CTG_UDABY76TF
shotgun clone IBEA_CTG_SZAAT45TF
shotgun clone IBEA_CTG_2007803
Sargasso1 (1)
cDNA clone PRNb47
cDNA clone PRNb61
cDNA clone PRNb67
cDNA clone PRNb16
Pelagibacter ubique ★
cDNA clone PRNb37
cDNA clone PRNb80
cDNA clone PRNb6
shotgun clone IBEA_CTG_1982036
shotgun clone IBEA_CTG_SKARN56TF
shotgun clone IBEA_CTG_UBAC706TR
cDNA clone PRNb89
shotgun clone IBEA_CTG_2151482
shotgun clone IBEA_CTG_SSAPM20TF
shotgun clone IBEA_CTG_UAAR518TF
cDNA clone PRNb63
cDNA clone PRNb44
shotgun clone IBEA_CTG_UBAZN75TF
shotgun clone IBEA_CTG_2154664
shotgun clone IBEA_CTG_2000453
shotgun clone IBEA_CTG_UHAAQ59TF
cDNA clone PRNb19
cDNA clone PRNb31
cDNA clone PRNb49
cDNA clone PRNb71
cDNA clone PRNb28
shotgun clone IBEA_CTG_1990461
shotgun clone IBEA_CTG_UEACY63TF
shotgun clone IBEA_CTG_SZAAO66TF
cDNA clone PRNb73
cDNA clone PRNb95
shotgun clone IBEA_CTG_1987166
shotgun clone IBEA_CTG_2075865

Taqman
Target

SAR11 and
SAR11-like

cDNA1 (27)   alpha
Group1 (1)
Sargasso2
EnvSe143, shotgun clone IBEA_CTG_1986497
MEDPR13, MEDPR13
13K9BAC, 13K9BAC
EnvSe148, shotgun clone IBEA_CTG_2157611
EnvSe152, shotgun clone IBEA_CTG_UDAX215TF
EnvSe161, shotgun clone IBEA_CTG_1983388
UncBact3, Uncultured bacterium clone medA15r8ex4 proteorhodopsin gene
UncBact4, Uncultured bacterium clone medA15r8b8 proteorhodopsin gene
UncBact5, Uncultured bacterium clone medA17_r8_11 proteorhodopsin gene
EnvSe138, shotgun clone IBEA_CTG_2038151
Sargasso3
cDNA2 (5)
EnvSe150, shotgun clone IBEA_CTG_UBAMN65TR
EnvSe172, shotgun clone IBEA_CTG_2103371
EnvSe160, shotgun clone IBEA_CTG_SZARE63TR
MEDPR46, MEDPR46
EnvSe159, shotgun clone IBEA_CTG_UDAPI94TF
EnvSe176, shotgun clone IBEA_CTG_UAANQ02TR
EnvSe147, shotgun clone IBEA_CTG_UDATA20TF
UncMarin, Uncultured marine alpha proteobacterium HOT2C01 fosmid clone.
EnvSe209, shotgun clone IBEA_CTG_2157217
EnvSeq40, shotgun clone IBEA_CTG_1986453
EnvSe163, shotgun clone IBEA_CTG_2000342
PRNb42, cDNA clone PRNb42
EnvSe169, shotgun clone IBEA_CTG_2156463
EnvSe174, shotgun clone IBEA_CTG_2073908
EnvSe193, shotgun clone IBEA_CTG_SSBWP28TF
EnvSe199, shotgun clone IBEA_CTG_2083842
EnvSe171, shotgun clone IBEA_CTG_2149707
EnvSe197, shotgun clone IBEA_CTG_2105235
EnvSeq50, shotgun clone IBEA_CTG_2145691
PRNb29, cDNA clone PRNb29
Group2

alpha

Green PR

EnvSe190, shotgun clone IBEA_CTG_2018987
SAR86-I (1)   gamma
PRNb26, cDNA clone PRNb26
ME42E12, MEDPR42E12, 328 bases, D849C5FD checksum.
EnvSe185, shotgun clone IBEA_CTG_SSBK845TF
EnvSe189, shotgun clone IBEA_CTG_UAAUB82TF
PRN03, cDNA clone PRN03
PRNb17, cDNA clone PRNb17
Sargasso4
EnvSe229, shotgun clone IBEA_CTG_2018343
EnvSe271, shotgun clone IBEA_CTG_SZAIB08TR
EnvSeq52, shotgun clone IBEA_CTG_2033912
EnvSeq51, shotgun clone IBEA_CTG_2111171
EnvSe208, shotgun clone IBEA_CTG_2151496
EnvSeq68, shotgun clone IBEA_CTG_2159084
Green PR
UncBac10, PCR clone medA15_r8ex7
UncBact7, PCR clone medA15_r9_3
UncBac14, PCR clone medA15r9b7
Blue PR
EnvSe237, shotgun clone IBEA_CTG_1997235
PRNb59, cDNA clone PRNb59
EnvSe249, shotgun clone IBEA_CTG_UAAQU03TF
EnvSe250, shotgun clone IBEA_CTG_2090075
ME35C06, MEDPR35C06, 324 bases, E62E49F5 checksum.
EnvSe255, shotgun clone IBEA_CTG_2043689
EnvSe353, environmental sequence
EnvSe245, shotgun clone IBEA_CTG_SKBMR59TF
Sargasso5 (1)
Sargasso6
GlbViol8, Gloeobacter violaceus PCC 7421 DNA
PyyLunu5, Pyrocystis lunula rhodopsin mRNA

gamma

gamma

0.10

Discussion

We were able to construct a PR cDNA clone library using RNA extracted from a bacterioplankton sample collected at the North Pacific central gyre station ALOHA and found that reverse transcription reactions using PR-specific degenerate primers were more efficient than those using random primers. These results somewhat contrast to those of Ståhlberg and coworkers who found that either random hexamers or oligo (dT) primers functioned most efficiently in the reverse transcription of five different genes (Ståhlberg et al. 2004a). One possible explanation for this discrepancy could be a differential effect of priming strategy for different genes, which was also suggested by the same study (Ståhlberg et al. 2004a) although random hexamers and oligo (dT) primers instead of specific primers alternated as the most efficient primer for different genes in that study. It is also possible that, in contrast to random hexamers that non-specifically target all mRNA molecules, specific primers anneal more efficiently when target mRNAs are in low abundance and in very complex mixtures, as was the case of PR mRNA in this study. Nevertheless, one apparent disadvantage of using target-specific primer in studies on multiple target genes is that separate RT reactions are required for each gene, imposing a higher demand for total RNA amount to be recovered from the samples.

We did not observe strong transcriptase effects on RT efficiency as previously suggested (up to 200-fold difference, Ståhlberg et al. 2004b), but we did observe that the use of the thermo-stable ThermoScript™ rather than the M-MLV RTase increased the specificity of RT reactions. Using the same reverse primer, RT reactions catalyzed with the M-MLV RTase resulted in the reverse transcription of non-specific

ribosomal RNA sequences, which did not occur with ThermoScript RTase. This is most likely attributable to the higher temperature (chosen as close to the Tm of the specific primer as possible) used in ThermoScript™ RTase catalyzed RT reactions that likely disrupted secondary structure of target mRNA molecules and reduced non-specific priming.

The high diversity of obtained PR cDNA sequences (Fig. 2.1) revealed that divergent bacteria carry PR genes and are also actively expressing these genes in the North Pacific Central Gyre. The large proportion of SAR11-like PR cDNAs in our cDNA clone library suggests that PR-harboring SAR11 or SAR11-like bacteria are most likely the predominant PR-expressing organisms at 25 m depth at Station ALOHA. We are aware that primer driven biases during RT or PCR might have affected these results, although the same primer set recovered different PR groups in the Mediterranean and Red Seas (Sabehi et al. 2005). This finding also indicates the potential for light driven energy contribution of SAR11 or SAR11-like bacteria, particularly when the fact that SAR11 bacteria were found to represent as much as 50% of the total surface water microbial community in the open ocean is taken into consideration (Morris et al. 2002). Interestingly, the PRs belonging to the Green-absorbing PR group that includes the originally identified PRs (Béjà et al. 2000a) was not found in the cDNA clone library constructed with primer set B, indicating that these organisms expressing Green-absorbing PRs may be less abundant in the open ocean, even though they were present. Green-absorbing PR cDNA sequences were the only types retrieved in the PR cDNA library constructed using the primer set A (data not shown), most likely because this primer set was designed based on a limited

number of sequences mostly belonging to this group (Sabehi et al. 2003). Finally, it is worth noting that despite the high diversity of PR sequences we have obtained so far, phylogenetic assignments were only possible for sequences related to five PR genes contained within large insert BAC or fosmid clones (Sabehi et al. 2005), whose phylogenetic origins were determined based on ribosomal RNA genes flanking these five PR gene sequences and the amino acid sequence of the cultivated *P. ubique*.

In addition to the widely used distance-based Neighbor-Joining method, we used the Bayesian method to infer phylogenetic relationships among PR proteins. For this inference, WAG was used as the substitution rate matrix instead of JTT, and the among-site substitution rate heterogeneity was corrected. Fairly similar PR phylogenies were generated using NJ and Bayesian methods. The most noteworthy discrepancy was the placement of the PR subclade we chose as target for TaqMan assays (Chapter 3) to the PR gene from the cultured SAR11 strain (Stapels et al. 2004). In the Bayesian tree, these two did not fall into the same clade as they did in the NJ tree, indicating that this PR subclade likely originated from a different SAR11 strain, or from other SAR11-like *Alphaproteobacteria*.

Since protein functionality and structure constraints evolutionary rates for different amino acid sites differ (Whelan et al. 2001) just as nucleotide substitution rates change with codon positions, we believed that the correction for among-site rate heterogeneity would avoid the otherwise possible underestimation of sequence distances due to the underestimation of substitutions occurring at fast-changing sites (Golding 1983; Jin and Nei 1990). Discrete gamma distribution was found to have sufficient performance to model site variation (Yang 1994), in which the α-parameter

decides the shape of the gamma distribution and thus the substitution rate divergence. The α-parameter estimated for our PR amino acid sequences was around 1, suggesting that most sites of the PR sequences (86 amino acids) from which the phylogeny was inferred had very low substitution rates or were virtually "invariable" (Yang 1996). However, this lack of rate variation is possibly due to the relative small number of amino acid positions rather than the reflection of the real substitution rates of complete PR proteins.

In conclusion, we were able to obtain a picture of the diversity of actively expressed PRs in the North Pacific Central Gyre, by constructing a PR cDNA clone library with degenerate primers. The same approach would be suitable for similar studies targeting other functional genes from uncultured microbes in the environment. More importantly, understanding the identity and diversity of actively-expressed genes in the environment would allow recognition of active members among microbes that harbor these specific genes, which, when combined with prior knowledge of different gene functions, can provide more insights into microbial roles in biogeochemical processes in marine ecosystems.

Literature cited

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN and DeLong EF (2000a). "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea." Science 289(5486): 1902-6.

Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, Villacorta R, Amjadi M, Garrigues C, Jovanovich SB, Feldman RA and DeLong EF (2000b). "Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage." Environ Microbiol 2(5): 516-29.

Béjà O, Spudich EN, Spudich JL, Leclerc M and DeLong EF (2001). "Proteorhodopsin phototrophy in the ocean." <u>Nature</u> 411(6839): 786-9.

Béjà O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T, Eisen JA, Fraser CM and DeLong EF (2002). "Unsuspected diversity among marine aerobic anoxygenic phototrophs." <u>Nature</u> 415(6872): 630-3.

Curry J, McHale C and Smith MT (2002). "Low Efficiency of the Moloney Murine Leukemia Virus Reverse Transcriptase during Reverse Transcription of Rare t(8;21) Fusion Gene Transcripts." <u>Biotechniques</u> 32(4): 768-775.

de la Torre JR, Christianson LM, Béjà O, Suzuki MT, Karl DM, Heidelberg J and DeLong EF (2003). "Proteorhodopsin genes are distributed among divergent marine bacterial taxa." <u>Proc Natl Acad Sci U S A</u> 100(22): 12830-5.

Fey A, Eichler S, Flavier S, Christen R, Hofle MG and Guzman CA (2004). "Establishment of a real-time PCR-based approach for accurate quantification of bacterial RNA targets in water, using Salmonella as a model organism." <u>Appl Environ Microbiol</u> 70(6): 3618-23.

Golding GB (1983). "Estimates of DNA and protein sequence divergence: an examination of some assumptions." <u>Mol Biol Evol</u> 1(1): 125-42.

Huelsenbeck JP, Ronquist F, Nielsen R and Bollback JP (2001). "Bayesian inference of phylogeny and its impact on evolutionary biology." <u>Science</u> 294(5550): 2310-4.

Jin L and Nei M (1990). "Limitations of the evolutionary parsimony method of phylogenetic analysis." <u>Mol Biol Evol</u> 7(1): 82-102.

Jones DT, Taylor WR and Thornton JM (1992). "The rapid generation of mutation data matrices from protein sequences." <u>Comput Appl Biosci</u> 8(3): 275-82.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A and Schleifer KH (2004). "ARB: a software environment for sequence data." <u>Nucleic Acids Res</u> 32(4): 1363-71.

Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA and Giovannoni SJ (2002). "SAR11 clade dominates ocean surface bacterioplankton communities." <u>Nature</u> 420(6917): 806-10.

Notredame C, Higgins DG and Heringa J (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." <u>J Mol Biol</u> 302(1): 205-17.

Ronquist F and Huelsenbeck JP (2003). "MrBayes 3: Bayesian phylogenetic inference under mixed models." <u>Bioinformatics</u> 19(12): 1572-4.

Sabehi G, Massana R, Bielawski JP, Rosenberg M, Delong EF and Béjà O (2003). "Novel Proteorhodopsin variants from the Mediterranean and Red Seas." <u>Environ Microbiol</u> 5(10): 842-9.

Sabehi G, Béjà O, Suzuki MT, Preston CM and DeLong EF (2004). "Different SAR86 subgroups harbour divergent proteorhodopsins." <u>Environ Microbiol</u> 6(9): 903-10.

Sabehi G, Loy A, Jung KH, Partha R, Spudich JL, Isaacson T, Hirschberg J, Wagner M and Béjà O (2005). "New Insights into Metabolic Properties of Marine Bacteria Encoding Proteorhodopsins." <u>PLoS Biol</u> 3(8): e273.

Sharkey FH, Banat IM and Marchant R (2004a). "Detection and quantification of gene expression in environmental bacteriology." <u>Appl Environ Microbiol</u> 70(7): 3795-806.

Sohail M and Southern EM (2000). "Hybridization of antisense reagents to RNA." <u>Curr Opin Mol Ther</u> 2(3): 264-71.

Stapels MD, Cho JC, Giovannoni SJ and Barofsky DF (2004). "Proteomic analysis of novel marine bacteria using MALDI and ESI mass spectrometry." <u>J Biomol Tech</u> 15(3): 191-8.

Ståhlberg A, Hakansson J, Xian X, Semb H and Kubista M (2004a). "Properties of the reverse transcription reaction in mRNA quantification." <u>Clin Chem</u> 50(3): 509-15.

Ståhlberg A, Kubista M and Pfaffl M (2004b). "Comparison of reverse transcriptases in gene expression analysis." <u>Clin Chem</u> 50(9): 1678-1680.

Suzuki MT, Béjà O, Taylor LT and Delong EF (2001a). "Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton." <u>Environ Microbiol</u> 3(5): 323-31.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H,

Pfannkoch C, Rogers YH and Smith HO (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." <u>Science</u> 304(5667): 66-74.

Wawrik B, Paul JH and Tabita FR (2002). "Real-time PCR quantification of rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes." <u>Appl Environ Microbiol</u> 68(8): 3771-9.

Whelan S and Goldman N (2001). "A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach." <u>Mol Biol Evol</u> 18(5): 691-9.

Whelan S, Lio P and Goldman N (2001). "Molecular phylogenetics: state-of-the-art methods for looking into the past." <u>Trends Genet</u> 17(5): 262-72.

Wilson MS, Bakermans C and Madsen EL (1999). "In situ, real-time catabolic gene expression: Extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater." <u>Appl Environ Microbiol</u> 65(1): 80-87.

Yang ZH (1994). "Maximum-Likelihood Phylogenetic Estimation from DNA-Sequences with Variable Rates over Sites - Approximate Methods." <u>Journal of Molecular Evolution</u> 39(3): 306-314.

Yang ZH (1996). "Among-site rate variation and its impact on phylogenetic analyses." <u>Trends in Ecology & Evolution</u> 11(9): 367-372.

CHAPTER 3: Development and Application of a Novel Method for the Measurement of Microbial Gene Expression in the Environment

Abstract

Transcriptional activity based on mRNA levels has traditionally been used as an indicator of expression levels of transcriptionally regulated genes. However, in studies involving environmental microbes, interpreting measurements of mRNA levels is complicated by the fact that many environmentally significant microorganisms are not represented by cultured strains. Consequently, mechanisms of gene expression regulation and even the linkage between metabolic genes and housekeeping or rRNA genes are often times not known. This lack of information, in turn, renders data normalization and interpretation difficult. Here, we developed a new method for the measurement of environmental microbial gene expression based on transcriptional activity, using the ratio between mRNA copies to gene copies as a measure of gene expression levels. Target mRNA and gene copies were measured in a "quasi-absolute" fashion (i.e., copy number per ml seawater assuming 100% efficiency of nucleic acid extraction and reverse transcription). This method was applied to study regulation and temporal dynamics of Proteorhodopsin (PR) gene expression at the 25 m at Station ALOHA in the oligotrophic North Pacific subtropical gyre, during three diel cycles. Proteorhodopsin mRNA levels were also normalized to total extracted RNA since total RNA has been previously used as a calibrator in gene expression studies and can compensate for variation due to possible RNA degradation. Similar diel patterns of PR mRNA expression were observed with both methods, except for some slight discrepancy in the beginning of the cruise, which was possibly caused by nucleic acid degradation. An increasing trend in PR transcriptional activity throughout the cruise observed when PR mRNA was

normalized to total RNA, but not when PR mRNA normalized to gene copies, was attributed to changes in microbial community composition (increased PR gene abundance). Although normalized PR mRNA levels peaked in the mornings, decreased in the afternoons and showed minima in midnight samples, suggesting possible light regulation of PR genes at the transcriptional level, fairly high PR gene expression was also observed after dusk, indicating light-independent effects. Hydrographic parameters revealed a strong internal tide signal at the 100 m depth, which shared synchronic dynamics to PR transcriptional levels at 25 m. Furthermore, fine-resolution density profiles showed isopycnal fluctuations around the 25 m isobath that helped explain fluctuations of PR gene copies in the seawater, as well as the late-night peaks of PR transcription.

Introduction

Currently, one of the primary goals of environmental microbiology studies is to expand from the current understanding of microbial community structure (i.e., what microbes live in the environment) towards the identification of the roles these microbes play in biogeochemical processes (i.e., what they are doing in the environment). Phylogenetic studies have shown that dominant bacterial phylotypes are poorly represented in culture collections (Giovannoni and Rappé 2000; Pace 1997), which is a chief factor precluding the thorough understanding of microbial physiology in laboratory settings. Recent developments in culture techniques allowed the cultivation of some strains belonging to previously uncultured bacterioplankton clades such as SAR11, and hence, provided further insights into microbial properties such as physiology, metabolism, and environmental significance (Kaeberlein et al.

2002; Rappé et al. 2002; Zengler et al. 2002). However, despite the remarkable achievements of cultivation-based techniques, the detection and measurement of many microbial activities in the nature still remains a challenge, mainly due to the following reasons: 1) the knowledge obtained from microbial physiology experiments conducted in laboratory does not necessarily reflect *in situ* microbial activities, and 2) for many bacterioplankton clades (e.g., SAR11, Roseobacter), strains characterized in pure culture represent only a small fraction of the total diversity and therefore may not be representative of environmentally significant strains.

Alternative cultivation-independent techniques have provided great resolving power in identifying and characterizing environmental microbial activities and allowed researchers to discover genes with known functions and monitor their expression levels in the environment. Since Pace and colleagues (Olsen et al. 1986) first applied molecular techniques in microbial ecology studies, culture-independent methods have undergone great advancements and created a more complete picture of microbial diversity and function in the ecosystem. More recently, the application of environmental genomics has resulted in the discovery of many novel genes from environmental samples (Béjà et al. 2000a; Stein et al. 1996), and larger-scale efforts such as sequencing of environmental "shotgun" libraries have further expanded our knowledge of the genetic diversity of bacterioplankton (Venter et al. 2004). Unlike PCR-based methods, prior knowledge of the sequence information of the target genes is not necessary to the construction of environmental genomic libraries, and thus it is possible to discover novel genes and to monitor specific microbial activities by the

measurement of expression levels genes associated with these activities in the environment.

mRNA levels have been broadly used in gene expression studies as surrogates for gene expression levels (Corredor et al. 2004; Wawrik et al. 2002; Wilson et al. 1999; Wyman et al. 1996). However, several potential factors may cause changes in mRNA levels that could be erroneously interpreted as variation in gene expression. Nucleic acid extraction efficiency might differ among samples, resulting in different amounts of starting material for quantification (Weinbauer et al. 2002). Reverse transcription (RT) reaction efficiency may also vary, leading to biases (Ståhlberg et al. 2004b). Finally, changes in mRNA signals in naturally occurring populations might also be caused by changes of cell numbers (gene abundance, Pichard and Paul 1993; Wyman 1999), or changes of bacterial community structure due to physical processes like advection, rather than transcriptional activity. Therefore, data normalization is crucial to the application of mRNA levels to estimate biologically relevant measurements of gene expression, particularly in environmental samples. Cell number (Wawrik et al. 2002), housekeeping genes (Dheda et al. 2004), ribosomal RNA (Zhong and Simons 1999) and total RNA (Bustin 2002; Leclerc et al. 2002) have been used as normalizers in eukaryotic gene expression studies. However, normalization approaches are limited for environmental gene expression measurements, since many environmentally significant species are not in pure culture and linkage between studied genes and rRNA genes or housekeeping genes are in most cases not known.

Here, we developed a novel method for the measurement of gene expression based on mRNA quantification, in which total nucleic acid extraction and RT efficiency were kept constant among samples, and possible variation in cell abundance was compensated by normalization to concentrations of the studied gene. In this new method, target gene and mRNA were quantitated using real-time quantitative PCR (Q-PCR) and reverse transcription PCR (Q-RT-PCR) (Gibson et al. 1996; Holland et al. 1991), respectively. Q-PCR allows DNA quantitation during the exponential phase of the reaction when the amplicon quantities are proportional to the starting template quantity, thus avoiding PCR biases associated with quantification at reaction end-points when amplicon quantities are not related to the template input due to PCR plateau effects (Suzuki and Giovannoni 1996). Compared with other mRNA quantitation assays such as Northern blot hybridization (Wyman et al. 2005), *in situ* hybridization (Parker and Barnes 1999), competitive PCR (Sharkey et al. 2004b), and RNase protection assays (Tomita et al. 2005), real-time RT PCR provides higher accuracy, sensitivity and wider detection range, and thus has become a widely used method for measurement of gene expression based on mRNA levels, and is especially suitable for samples with low mRNA abundance (Walker 2002). Various detection chemistries are currently available for Q-PCR, and here we used TaqMan assays (also known as 5'-nuclease assays).

We attempted to measure transcriptional activity of the light-driven proton pump Proteorhodopsin genes *in situ* to better understand of PR gene expression regulatory mechanisms. As discussed in the introduction section of this thesis, Proteorhodopsins (PRs), like archaeal bacteriorhodopsins (BRs), function as light-

driven proton pumps to produce cellular energy. Studies by Shand and Betlach (Shand and Betlach 1991) have shown that expression the of bacteriorhodopsin gene of *Halobacterium halobium* is induced by light and low oxygen tension. Transcriptional level regulation of PR genes has also been suggested in studies by Baliga and co-workers (Baliga et al. 2001) based on the existence of regulatory sequence regions associated with the PR genes of the *Gammaproteobacterium* SAR86. Since oxygen levels are fairly high and constant at the surface in oligotrophic ocean waters, we hypothesized that light is the main environmental regulator for PR gene expression at station ALOHA in the North Pacific Central Gyre.

Here, we provide evidence supporting the hypothesis that PR gene expression is regulated at the transcriptional level. Furthermore, the fluctuation of measured PR gene expression throughout the diel cycles in the North Pacific subtropical gyre suggests not only light regulation but also effects of physical processes on the measured PR gene expression. This novel approach for the measurement of *in situ* gene expression can potentially be used in future studies of microbial gene expression in the environment.

Materials and methods

Sample collection and nucleic acid extraction.

Seawater was collected from 25 m depth at the Hawaii Ocean Time Series station ALOHA (22º44´N, 158º2´W) at 21 times points (approximately 3 hours apart) from 12:00 AM on December 13, 2002 to 12:00 AM on December 16, 2002.

For RNA extraction, bacterioplankton from about 3800 ml of seawater was collected onto a 25 mm 0.2 μm polysulfone filter (Pall Gelman Inc.), frozen and kept at –20ºC aboard the R/V Kilo Moana before shipped frozen to Solomons where it was stored at –70ºC until RNA and DNA extraction procedures. Total RNA extraction and DNA removal were performed following the procedures described in Chapter 1.

For DNA extraction, bacterioplankton samples from about 400 ml seawater were collected onto 13 mm Supor200® 0.2 μm filters (Pall Gelman) and DNA extracted following previously described procedures (Suzuki et al. 2001b).

RNA and DNA quantification.

RNA concentration was measured for all 21 samples exactly as described in Chapter 1. DNA samples were quantified by PicoGreen® (Molecular Probes) staining following the same procedures as in RNA quantification except that a Typhoon™ 9410 fluorescence imager (Amersham Biosciences, Piscataway, NJ) instead of the Spectra MAX Gemini microplate spectrofluorometer was used, due to the relatively low DNA concentration in DNA extracts.

Primers and probe design.

The primers and TaqMan probe targeting the most prevalent PR group in a PR cDNA clone library (Chapter 2), which was affiliated with *Pelagibacter ubique* (SAR11) clade. The reverse primer was used in reverse transcription reactions to produce the target cDNA (Table 3.1). The TaqMan probe was designed to anneal to the anti-sense strand of the templates, and was labeled with the reporter dye 6-carboxyfluorescein (6'-FAM) at the 5' end and with the quencher dye 6-carboxytetramethylrodamine (TAMRA) at the 3' end. The primers and probe were

designed with ARB software (Ludwig et al. 2004), using an alignment of all available

PR genes in GenBank (as of July, 2003). The Primer Express$^©$ software (Applied

Biosystems, Foster City, CA) was used to check primer properties such as T_m and

possible secondary structure.

Table 3.1. Primers, probe, enzymes and reaction parameters used in real-time RT-PCR.

RT	**Specific primer:**	5´-CAA GTG TAG CGT TGA TGT AAC C-3´ (final conc. 1μM)
	Enzyme:	ThermoScript™ RNase H− Reverse Transcriptase (Invitrogen)
	Parameters:	52 ºC, 55 min, 85 ºC, 5 min
Taqman	**Forward primer:**	5´-GCG ATA AGA AAA GTA CCT TCA GG-3´ (final conc. 1,000 nM)
	Reverse primer:	5´-CAA GTG TAG CGT TGA TGT AAC C-3´ (final conc. 500 nM)
	Probe:	5´-TCC ACT TAC TAG CAT TAC TAC TGA TCC GA-3´ (final conc. 200 nM)
	Enzyme:	Platinum® Taq DNA polymerase (Invitrogen)
	Cycle parameters:	50 ºC, 2 min; 95 ºC, 2 min; 40 cycles (95 ºC, 15 sec; 58 ºC, 1 min)

Plasmid standards.

We used recombinant cDNA as standards for TaqMan assays. PR cDNA

cloning procedure was described in Chapter 2 and the plasmids from clone PRNb28

were extracted, linearized and purified, using a modification of a previously described

protocol (Suzuki et al. 2000). Briefly, plasmids were purified using Eppendorf$^®$

FastPlasmid™ Mini kit (Eppendorf, Westbury, NY) according to the manufacturer's

protocol and further treated with plasmid-safe™ ATP-dependent DNase (Epicentre,

Madison, WI) to remove contamination of *E. coli* chromosomal DNA. 10 µl of

plasmids were digested with the restriction endonuclease *Not*I (Promega, Madison

WI) according to the manufacturer's protocol, purified by phenol-chloroform treatment followed by dialysis in spin columns (Microcon 100, Millipore, Billerica, MA) and subjected to three washes with TE buffer using Micro Bio-Spin Chromatography columns (Bio-Rad, Hercules, CA). Copy numbers of this plasmid standard was measured via TaqMan assays using the pUC primer and probe set, using plasmid EBAC31A08 as a "gold standard" (Suzuki et al. 2001b).

TaqMan assay optimization.

Primer Specificity. Since the primers and probe were designed to target a specific set of PR genes, a specificity test was performed as follows. Plasmid concentrations of 96 clones in the PR cDNA clone library (Chapter 2) were measured spectrophotometrically using a SpectraMax Plus[384] spectrophotometer (Molecular Devices, Sunnyvale, CA), and subsequently diluted to 0.1 ng/µl with nuclease-free TE buffer (Ambion, Austin, TX). Ten target clones and 31 non-target clones most likely subject to non-specific primer annealing were used as template in PCR reactions. Ten µl PCR reactions contained 1 µl of PCR 10X buffer (Invitrogen, Carlsbad, CA), 0.2 mM of dNTPs (Promega, Madison, WI), 3 mM of $MgCl_2$ (Invitrogen), 0.5 µM of primers, and 0.05 U µl^{-1} of Platinum® Taq DNA polymerase (Invitrogen). The cycle parameters were as follows: 2 min at 92°C and 35 cycles of 30 sec at 92°C, 30 sec at 58°C, and 30 sec at 72°C. PCR products from entire reactions were loaded in a 2.5% Nusieve® GTG® agarose Gel (Cambrex, Rockland, ME), which was electrophoresed at 80 V for 2 h and post-stained with 1:10,000 SybrGreen I (FMC Bioproducts, Rockland, ME) for 20 min.

Primer concentration. A primer concentration matrix experiment was performed seeking the lowest C_Ts in TaqMan assays. In each of the 25-µl TaqMan assay reactions, reagent concentrations except for primer concentrations were kept constant: 2.5 µl 1X Platinum® Taq DNA polymerase buffer, 0.2 mM of dATP, dCTP, and dGTP, 0.4 mM of dUTP, 5mM of $MgCl_2$, 200 nM of TaqMan probe, 0.01 U µl^{-1} AmpErase® Uracil N-Glycosylase (UNG, Applied Biosystems), 1.2 µM of 5'-labeled ROX reference dye (Megabases Inc, Evanston IL), 0.025 U µl^{-1} of Platinum® Taq DNA polymerase (Invitrogen). A matrix of different concentrations of forward and reverse primers ranging from 100 nM to 1500 nM (100 nM, 500 nM, 1000 nM, and 1500 nM) was tested. 22.5 µl of the master mix were distributed into Optical 96-Well Reaction Plates (Applied Biosystems) using a Microman® M100 positive displacement pipette (Rainin, Emeryville, CA), followed by the addition of 2.5 µl of the PR standard (concentration: 4.5 x 10^5 copies/µl). All reactions were performed in an ABI Prism SDS 7000 system (Applied Biosystems) and cycle parameters are listed in Table 3.1. The concentrations of forward and reverse primers yielding the highest ΔRn and lowest C_T were chosen for optimization of probe concentration.

Probe concentration. Five TaqMan reactions with probe concentrations varying from 50 nM to 250 nM (50 nM, 100nM, 150nM. 200nM, and 250nM) were performed to seek for the probe concentration that would give the highest amplification efficiency. In 25 µl final volume, forward primer concentration was 1000 nM, reverse primer concentration was 500 nM, and concentrations of other reagent were the same as described above in the *primer concentration* subsection

above. All reactions were performed in an ABI Prism SDS 7000 system (Applied Biosystems) and cycle parameters listed in Table 3.1.

Passive reference dye.

In addition to the 5'-labeled ROX dye, SuperROX™ (Biosearch Technilogies, Novato, CA) was tested as passive reference dye and shown to produce higher fluorescence signal and smoother amplification curves in the TaqMan assays. Therefore, SuperROX™ dye was used in all subsequent TaqMan assays.

Measurement of gene expression.

We used different approaches for measuring gene expression levels based on gene transcription. The novel approach developed here (per gene normalization, hereafter) used the ratio between PR mRNA copies per ml seawater to PR gene copies per ml seawater to estimate gene transcription, with PR mRNA and gene abundance measured in a "quasi-absolute" fashion, assuming 100% extraction efficiency and 100% RT efficiency. This novel approach was compared to one previous approach where total RNA was used as the normalizer of gene transcript abundance to account for possible variation due to nucleic acid degradation (Bustin 2002; Leclerc et al. 2002). Finally, we also used the ratio of PR mRNA per total RNA to PR gene per total DNA as an additional measure of gene expression that could compensate for target gene abundance variation.

rRNA gene quantification.

Bacteria and SAR11 16S rRNA gene abundances were measured using different TaqMan assays as previously described (Suzuki et al. 2001b).

Statistical analysis.

To analyze the diel patterns of PR gene expression, and in particular, to validate the bimodalities indicated by both methods used in this study, we conducted harmonic regression analysis using SAS 8.0.2 (SAS Institute Inc. 1985). Harmonic regression (also known as trigonometric regression or cosinor regression) is a linear regression model in which the predictor variables are trigonometric functions of a single variable, usually a time-related variable. It is always used to model biological signals that tend to have stable periodic rhythms, such as diurnal, circadian or yearly rhythms. We used least-square techniques to obtain parameter estimates of the equation (Miller et al. 1995):

$$Y_{jt} = \beta_0 + \mu_j + \sum(\beta_{1k}\sin(k\omega t) + \beta_{2k}\cos(k\omega t)) + \varepsilon, \text{ where } k=1, 2, \ldots n.$$

In this equation, Y_{jt} is the measured PR gene expression level; μ_j represents the jth day effect; β_{1k} and β_{2k} are estimated parameters for a given k value; ω is the frequency expressed in terms of radians per unit time, that is: $2\pi/24$, where π is the constant *pi* and 24 is the frequency of a diel cycle (24 h); the variable t is a continuous numeric values converted from time variable (e.g., 2:24 AM in day 2 would be converted to 26.4). Significant first-order terms, i.e., k=1, indicate a dome-shaped diel pattern, and second-order terms (k=2) indicate a bimodal diel pattern, and so on. In addition, we also explored the day effect, as represented by μ_j, to the regression model by creating 3 dummy variables: day1, day2 and day3 (day1=1 for the expression levels measured on the first day, otherwise day1=0; the same for variables day2 and day3).

Results

TaqMan primer and probe set.

The sequences of the TaqMan primers and probe are shown in Table 3.1. The calculated T_m (melting temperature) at 500 nM concentration according to Primer Express$^©$ software was 58 °C for the forward primer, 57.4 °C for the reverse primer, and 63 °C for the TaqMan probe. The end-point PCR specificity test for TaqMan primers showed that all the target clones tested gave positive PCR products detected by NuSieve$^®$ gel electrophoresis, while all but one of 31 non-target clones yielded none or very little PCR products at 58 °C (Fig. 3.1). The considerable non-specific amplification of clone PRNb95 (showed in red) was checked by using 1 µl of plasmid prepared from a freezer stock of PRNb95 clone (in LB media plus 7% Glycerol) as PCR template and shown to result from cross contamination during the plasmid preparation. In addition to the high specificity of the primer set, the TaqMan probe designed with no degeneracy to specifically target the PR group provided further assurance of the specificity of TaqMan assays.
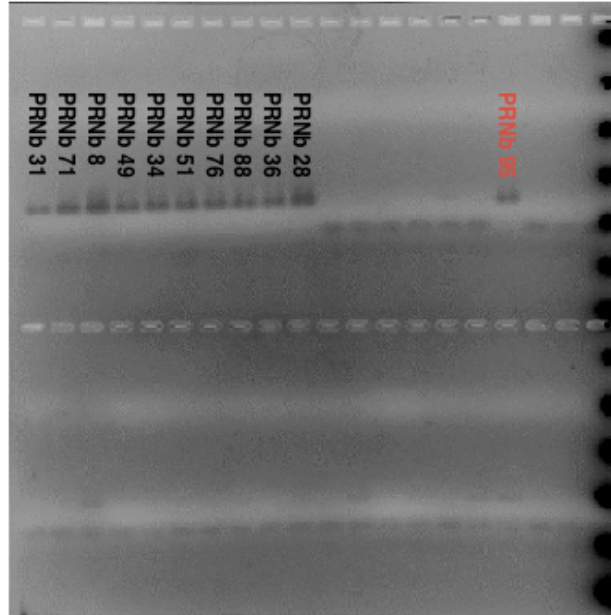
Fig. 3.1. The end-point PCR test for primer specificity. The lanes marked with clone names in black were loaded with PCR products of target clones. The remaining lanes were loaded with PCR products of non-specific clones, with one showing positive result (marked in red).

TaqMan assay optimization.

As shown in figure 3.2, the combination of 1000 nM forward primer and 500 nM reverse primer yielded the highest ΔRn and lowest C_T, which indicated highest amplification efficiency. 200 nM was shown to be the optimized final concentration of the probe. Different SuperROX™ concentrations were tested and a final concentration of 1.2 μM was found to give a similar fluorescence to TaqMan probe reporter dye (FAM) in the first cycle of the amplification, and yielded smooth amplification curves.
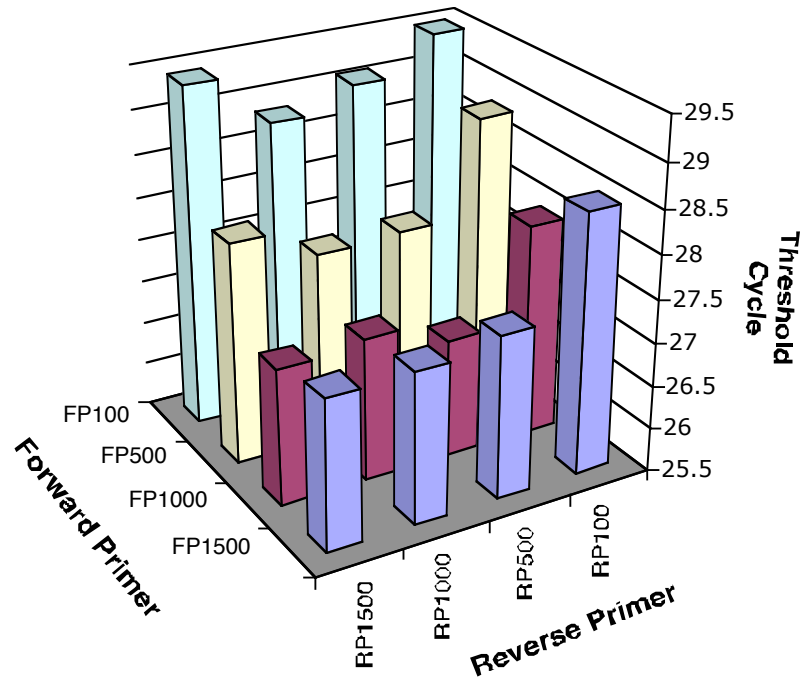
Fig. 3.2. Primer optimization matrix experiment.

Standard curves.

Plasmids containing cloned cDNA were used as standards in TaqMan assays quantifying both PR mRNA and gene copy numbers. The standard curves showed a strong linear correlation ($R^2 >0.99$) between C_T and the log of starting copy number over 4 orders of magnitude starting from 400 copies μl^{-1}. The slopes of standard curves were ranged between -3.28 to -3.32, which, according to the equation: $E=10^{[-1/slope]}$, indicated high amplification efficiencies (near 100%). The use of recombinant DNA as standard for mRNA quantification was validated by the mRNA linear extraction experiment described in details in Chapter 1.

Measurement of PR gene transcription.

Figure 3.3 shows concentrations of PR mRNA and its encoding gene (copies/ml) in the seawater samples during the cruise. The regression analysis

indicated a significant trend of increasing PR mRNA and gene abundance with time

(Fig. 3.3A, Pearson r=0.63, p=0.001; Fig. 3.3B, Pearson r=0.66, p<0.001).
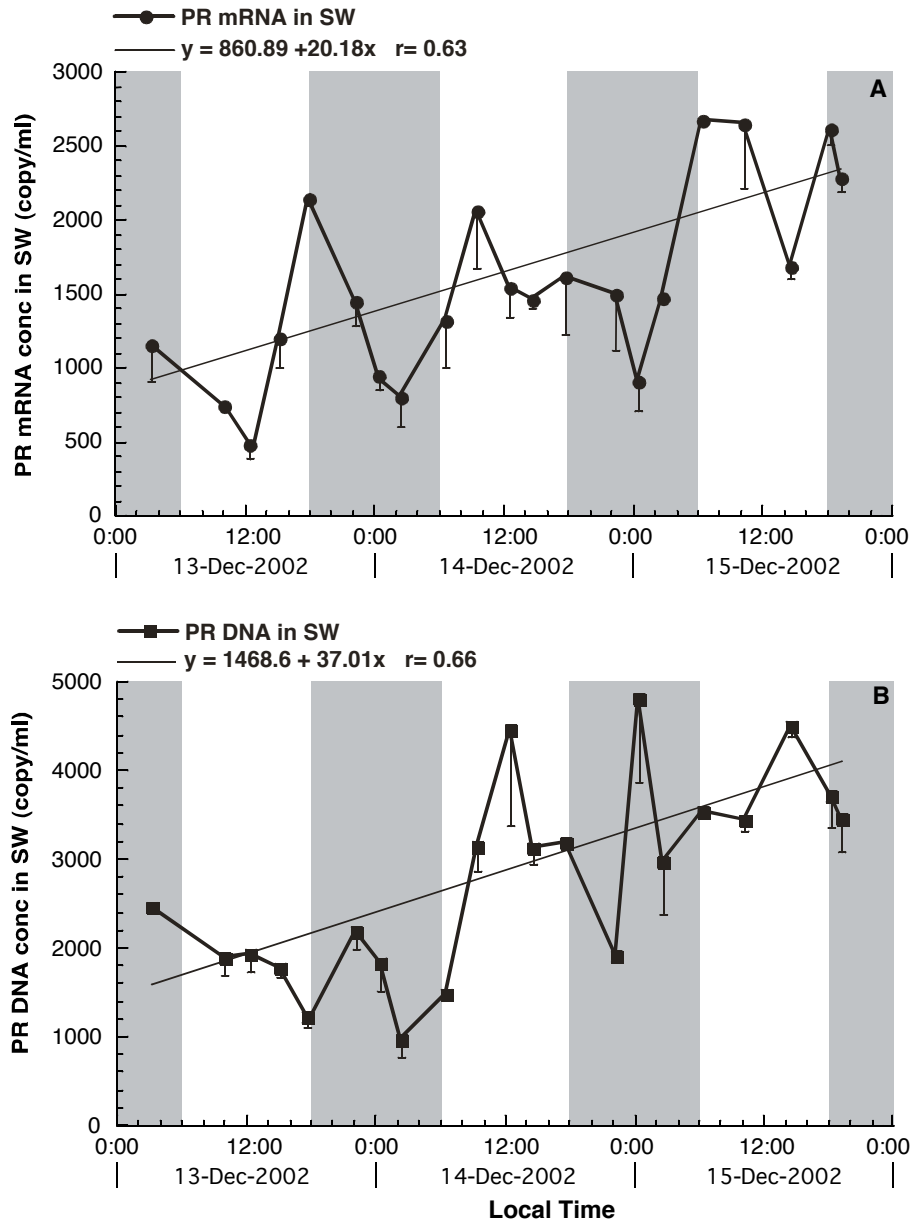
Fig. 3.3. PR mRNA **(A)** and gene **(B)** concentrations (copy/ml) in seawater samples during the cruise, measured in a quasi-absolute fashion. Error bars represent one standard deviation of repeated measures (n=3). Local times were converted to continuous numeric values (e.g., 2:24 am in day 2 would be converted to 26.4) when subjected to regression analysis with y values. This time conversion was performed for all regression analyses involving local time in this study.

Figure 3.4 illustrates the diel dynamics of PR transcript levels measured using per gene normalization method during the 3-day cruise. As shown in figure 3.4A, PR transcriptional activity peaked in early mornings and showed minimum values in midnight samples. We also observed consistent decreases of PR transcription from early mornings to late afternoons and subsequent increases from late afternoons to late nights (around 10 PM). However, the regression analysis (slope=0.00078, Pearson r=0.0079) indicated that, except for the diel variations, the expression of PR genes did not show a trend of general increase with time as seen with PR mRNA and PR gene.

Similar diel patterns of gene transcription were observed when mRNA levels were normalized to total RNA (Fig. 3.4B), except for a modest trend of general increase in PR transcriptional activity throughout the cruise revealed by regression analysis (Pearson r= 0.43, p=0.05). Similar to PR mRNA copies per total RNA, PR gene copies per total DNA also increased with time (Fig. 3.4C). Finally, unlike PR mRNA copies per total RNA or PR gene copies per total DNA, the ratio between these parameters (PR mRNA normalized to total RNA over PR DNA normalized to total DNA) did not show an increasing trend with time, but it shared similar dynamic patterns as PR transcript levels measured by the per gene normalization (Fig. 3.4D), except for some discrepant values measured in the first day of the cruise, possibly resulting from nucleic acid degradation. Regression analysis of PR gene transcriptional levels for the samples taken in the last two days of the cruise revealed significant correlation between the two measurements (Pearson r=0.65, p=0.0048).
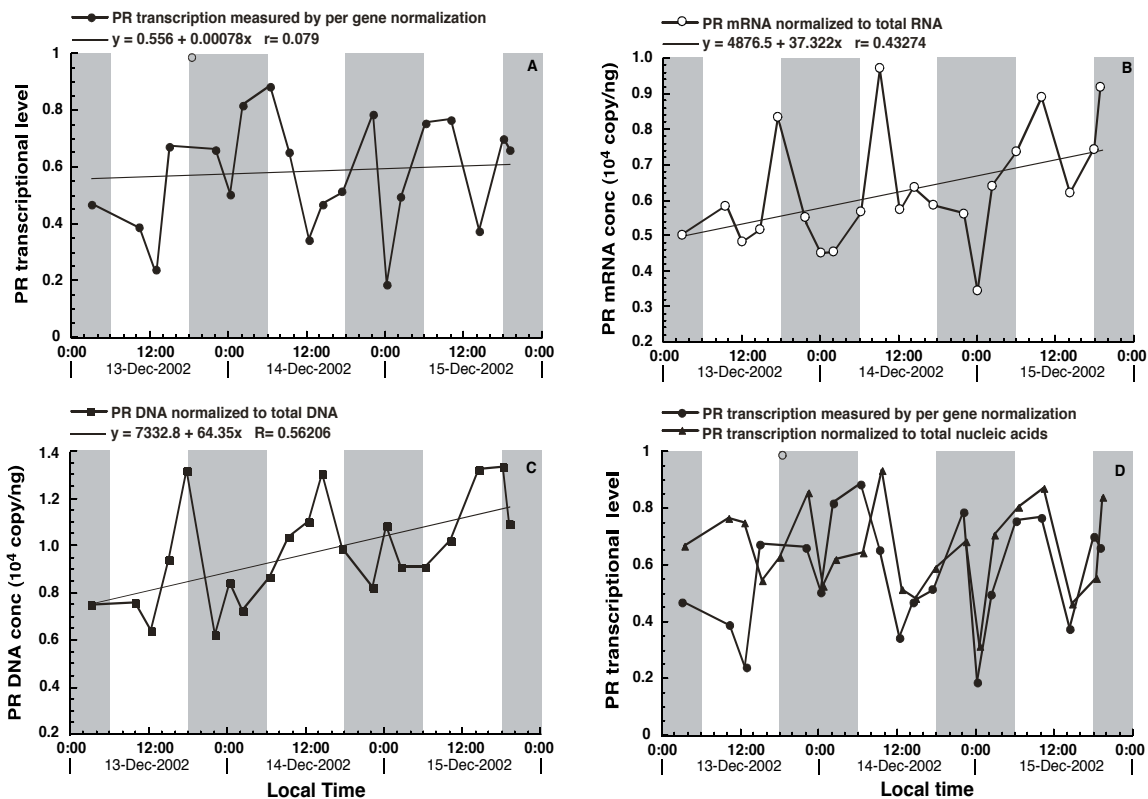
Fig. 3.4. (A) PR transcriptional activity measured by using the ratio of PR mRNA concentration in the seawater normalized to PR gene concentration in the seawater (per gene normalization). The open circle represents a possible outlier. **(B)** PR mRNA concentration in the total extracted RNA. **(C)** PR gene concentration in the total extracted DNA. **(D)** Comparison of PR transcriptional activity measured by using two methods: per gene normalization, and PR mRNA concentration in the total RNA over PR gene concentration in the total DNA.

Harmonic regression of PR transcriptional levels.

The regression fit of Harmonic Regression Analysis is shown in Figure 3.5. No significant day effect on the PR gene transcription was shown among the three days (p-values for dummy variables day1 and day2 are 0.40 and 0.84, respectively; data not shown), which suggested that except for the diel patterns, PR gene transcriptional levels show no significant day-to-day variation. Therefore, we excluded day effect variable μ_j from our model.

The Harmonic regression model indicated a significant bimodal diel pattern for PR gene transcription levels (p=0.019) (Table 3.2). Morning peaks and late-night peaks, as well as mid-night minimums for PR gene transcription were predicted. However, the p-value was slightly greater than 0.05 (Table 3.2), indicating effects other than diel regulation on measured PR gene transcription.

Table 3.2. Harmonic regression parameter estimates of the PR transcription measured using the per gene normalization method.

PR expression	$R^2 = 0.46$, DF $= 19$, p $= 0.066$			
Parameter	Estimates	SE	t value	p
Intercept, β_0	0.600	0.038	15.51	<0.0001
1st order sine, β_{11}	0.051	0.057	0.89	0.39
1st order cosine, β_{21}	0.046	0.051	0.91	0.38
2nd order sine, β_{12}	-0.084	0.053	-1.57	0.137
2nd order cosine, β_{22}	-0.144	0.055	-2.62	**0.019**
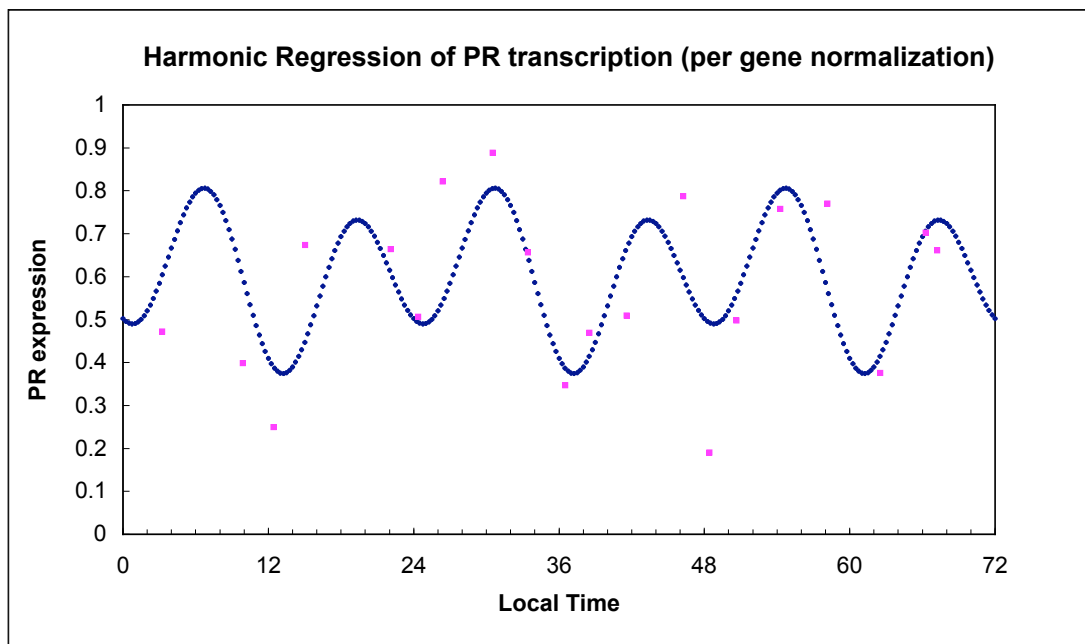
Fig. 3.5. Harmonic Regression of PR gene expression levels measure using the novel approach (per gene normalization). The dotted line is the regression fit, and the squares in red are measured PR transcriptional levels.

Effects of physical processes on measured PR gene transcriptional levels.

CTD isohaline contours indicate that a high salinity water intrusion started around 12:00 AM on 14 December, 2002 (Fig. 3.6) and likely caused the bacterial community to change as indicated by the increase in nucleic acid concentrations (Fig. 3.8B), PR mRNA over total RNA (Fig. 3.4B), and PR gene over total DNA (Fig. 3.4C). In order to evaluate the changes in bacterioplankton community with the high salinity intrusion, we also measured total *Bacteria* and SAR11 rRNA gene abundance (Fig. 3.7). *Bacteria* rRNA gene abundance increased significantly between the first (prior to 12:00 PM on December 14, 2002) and the second half (post 12:00 PM on December 14, 2002) of the cruise (student's t, p=0.003). The PR genes, which we believe are likely harbored by a bacterium in the SAR11 clade (Chapter 2), showed

an increasing percentage over SAR11 rRNA genes (Pearson r=0.63, p=0.002). Even

though, the abundance of the PR genes and SAR11 rRNA genes showed a good

correlation (Pearson r=0.76, p=0.0004).
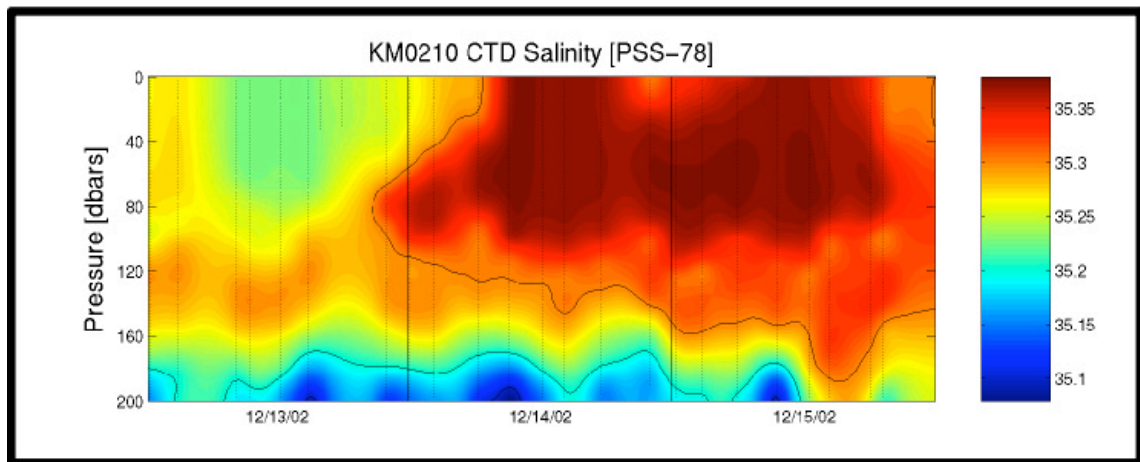
Fig. 3.6. Cruise CTD isohaline contours.

Also evident in the isohaline contours (Fig. 3.6) was the internal tide with 12 h

frequency (Margaret A. McManus, University of Hawaii, Personal communication)

clearly seen at about 100 m (Fig. 3.6). However, the intrusion of the high salinity

water mass masked the tidal signal at shallower depths, in particular at 25 m where

the samples used to measure PR gene expression were collected. Fortunately,

isopycnal contours at a higher resolution enabled visualization of physical features

after the intrusion (Fig. 3.8A), suggesting that 25 m depth water layer was also

affected by internal tides, though with a shorter frequency (10.5 h, Margaret A.

McManus, University of Hawaii, Personal Communication).
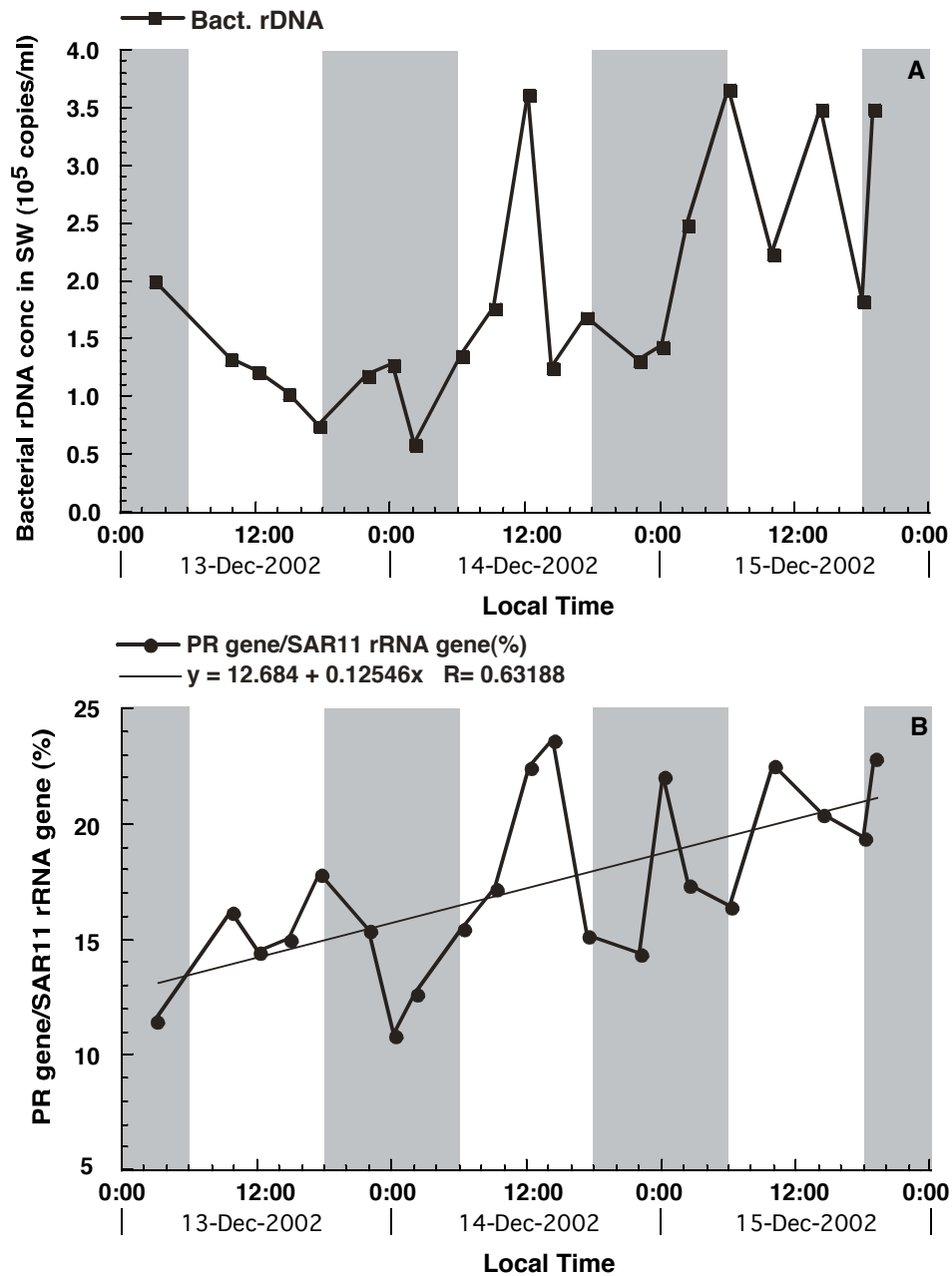
Fig. 3.7. **(A)** Total bacterial rRNA gene abundance measured by TaqMan assays. **(B)** The ratio between PR gene abundance to SAR11 rRNA gene abundance.

A comparison between isopycnal contours (Fig. 3.8A) and total nucleic acid concentrations at 25 m (Fig. 3.8B) show a high correlation between isopycnal compressions and total nucleic acid concentration increases, indicated by the vertical

bars. Additionally, isopycnal relaxations and total nucleic acid decreases were also highly correlated (Fig. 3.8).
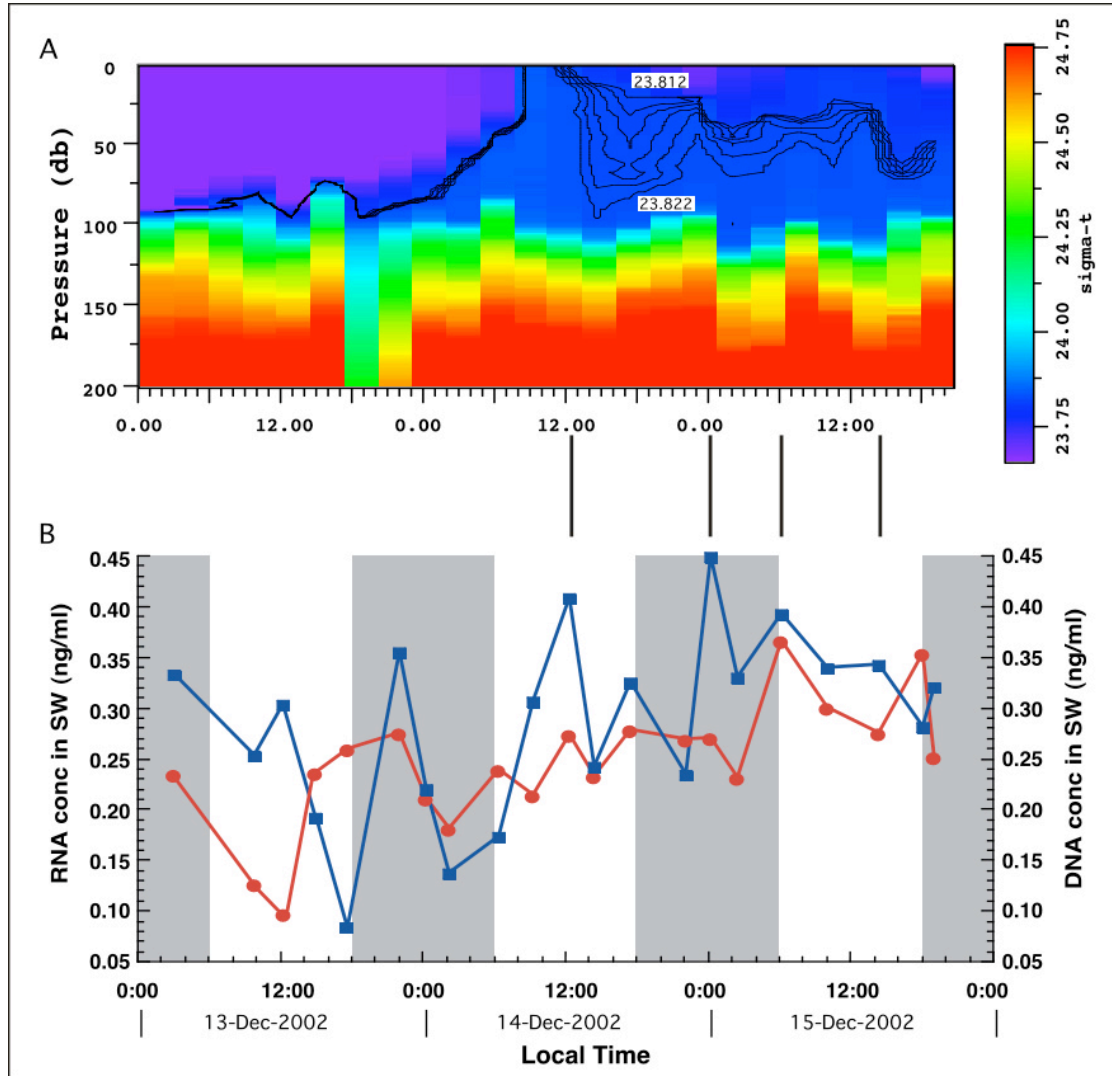
Fig. 3.8. (A) Cruise CTD isopycnal contour figure. **(B)** Concentration of total nucleic acids. The four vertical bars in between panel A and B indicates times when both isopycnal compressions and total nucleic acid concentration peaks occurred. Blue: DNA (ng/ml). Red: RNA (ng/ml).

Similar relationships were observed for isopycnal contours and PR gene concentrations at 25 m. A notable exception occurred at 10 PM on December 14,

2002, when a very low concentration of PR gene was measured shortly before a strong isopycnal compression event, and very quickly increased to a high value at the time of the isopycnal compression (Fig. 3.9).
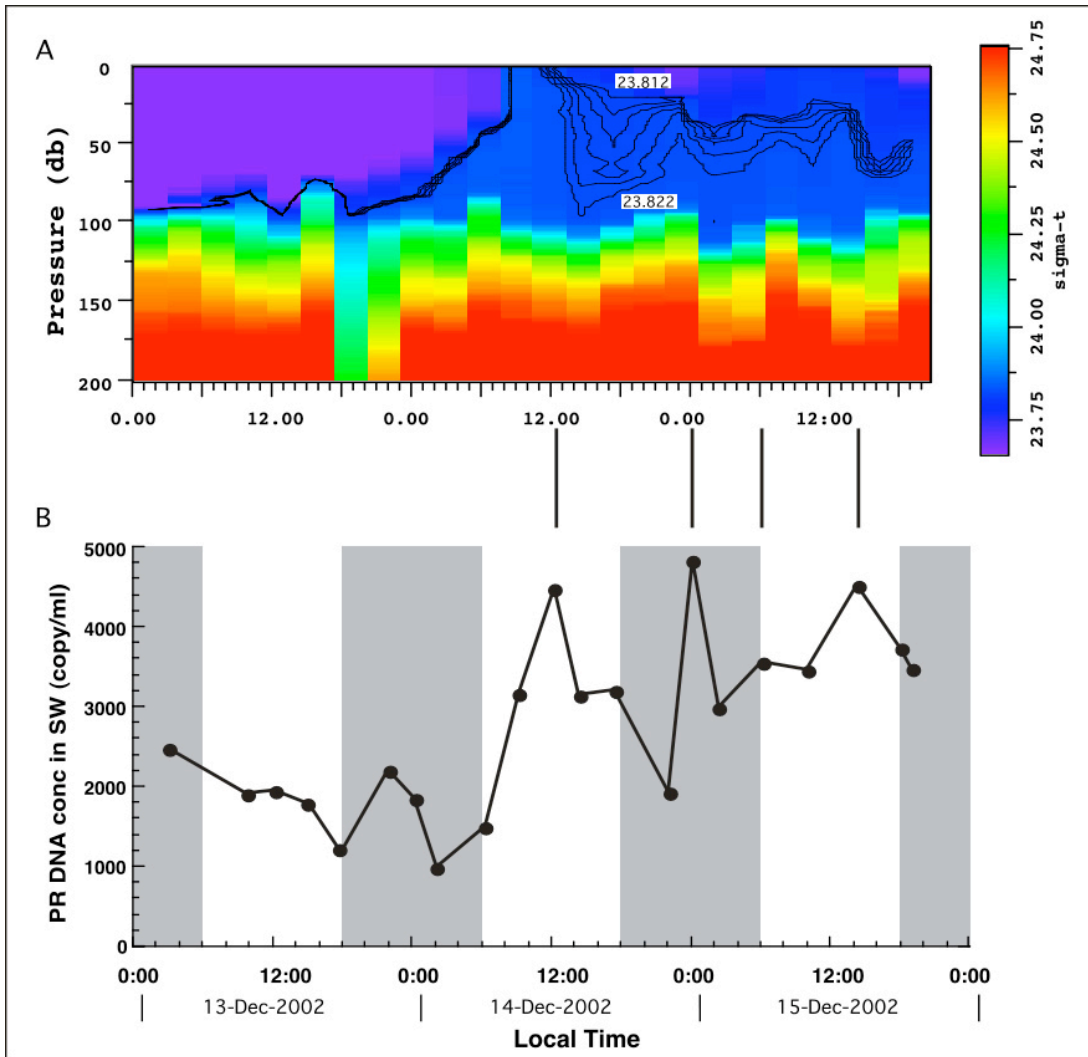
Fig. 3.9. **(A)** Cruise CTD isopycnal contour figure. **(B)** Concentrations of PR genes in the seawater. The four vertical bars in between panel A and B indicates times when both isopycnal compressions and PR gene concentration peaks occurred.

A closer look at the enlarged isopycnal contours and overlaid PR gene concentrations (Fig. 3.10) revealed that PR gene concentrations were correlated with

the upward and downward fluctuation of an isopycnal that periodically crossed 25 m isobath (σ_θ=23.812 ± 0.002 kg.m^{-3}), likely resulting from the internal tides.   In general, PR gene concentrations decreased with a shallowing of the isopycnal, and increased with the deepening of this isopycnal. The unexpectedly low concentration of PR gene concentration mentioned above (Fig. 3.9), corresponded to an apparent shallowing of this isopycnal (steep upward intrusion of deeper water) (Fig. 3.10).
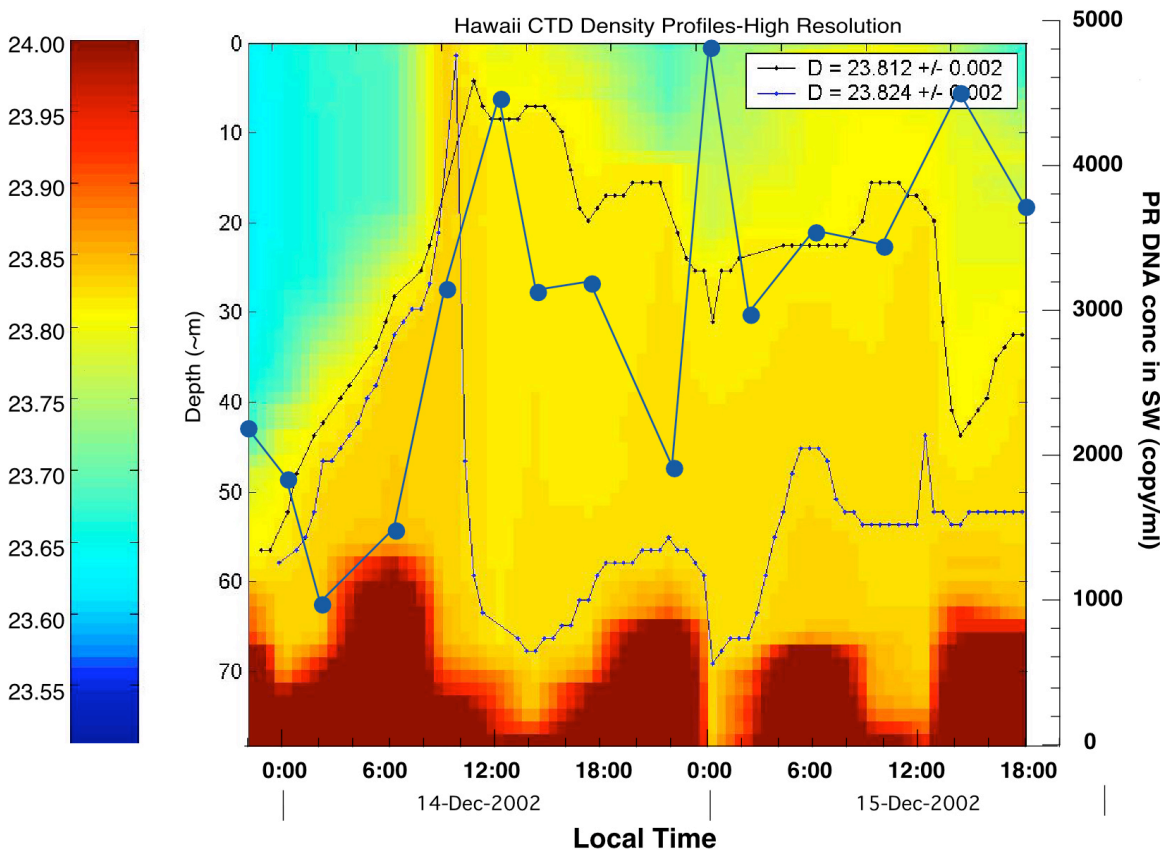
Fig. 3.10. An enlargement of isopycnal contours for the second two days of the cruise, with PR gene concentration in the seawater overlaid on top. The two isopycnals represent σ_θ=23.812 ± 0.002 kg.m^{-3}, and σ_θ=23.824 ± 0.002 kg.m^{-3}, respectively.

Figure 3.11 shows a comparison between isopycnal contours and PR mRNA concentrations in the seawater. Unlike PR DNA concentrations, PR mRNA concentrations did not always increase with isopycnal compressions, and in fact, three out of four isopycnal compressions were associated with decreases in PR mRNA concentrations, indicating low PR transcription levels. The single PR mRNA concentration that agreed with the isopycnal compression was measured in the early morning of December 15, 2002.
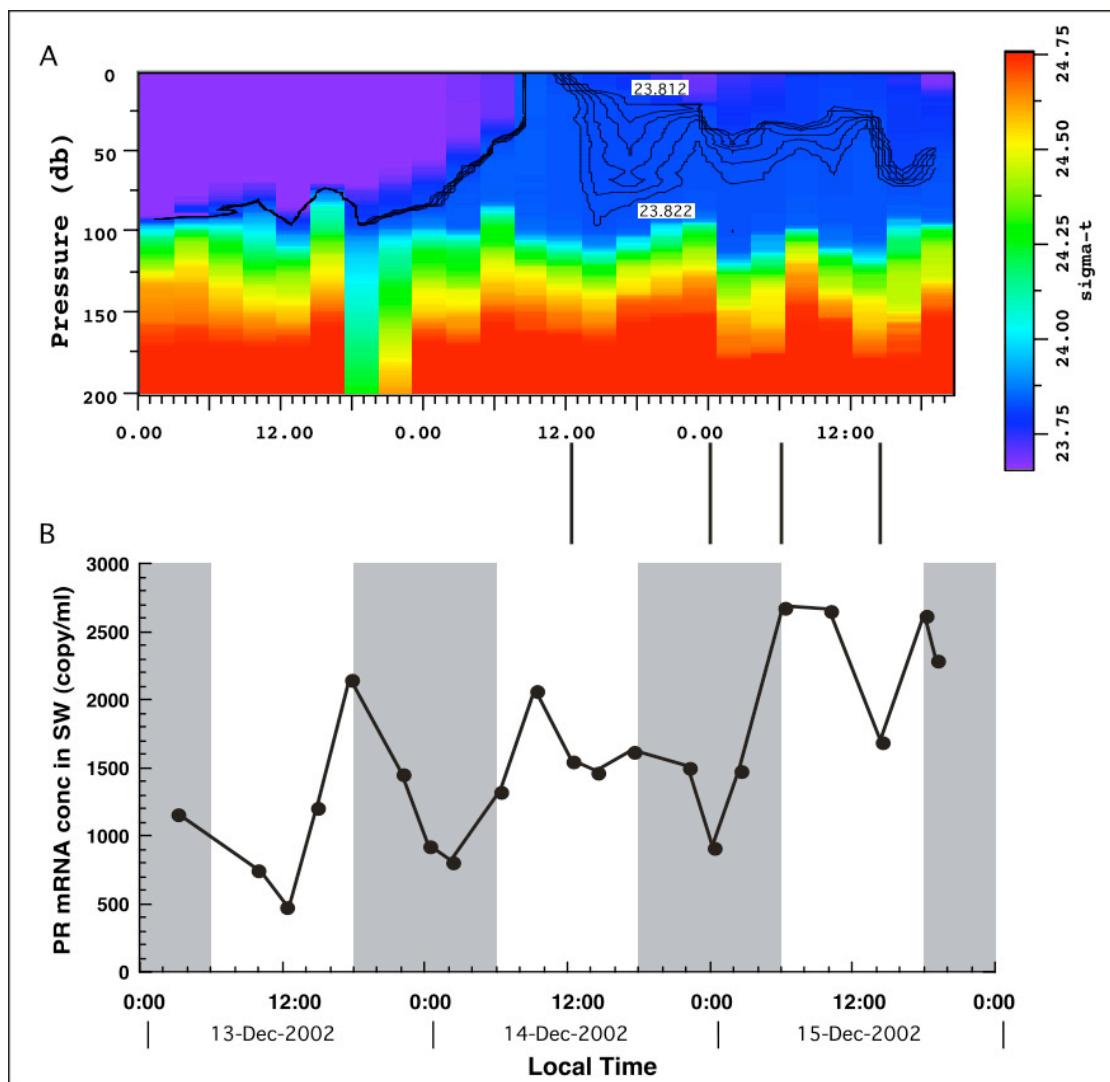
Fig. 3.11. **(A)** Cruise CTD isopycnal contour figure. **(B)** Concentrations of PR mRNAs in the seawater. The four vertical bars in between panel A and B indicates times when isopycnal compressions occurred.

PR mRNA concentrations were also correlated with fluctuations of same the isopycnal (σ_θ=23.812 ± 0.002 kg.m^{-3}), but in opposite fashion to that of PR genes: PR mRNA concentrations increased with a shallowing of the isopycnal, and decreased with the deepening of the isopycnal (Fig. 3.12). Two apparent exceptions (5:30 PM, December 14, 2002, and 10:10 AM, December 15, 2002) were observed, when a

deepening of the isopycnal was not associated with decreases of PR mRNA concentrations. Instead of decreases, PR mRNA concentration increased or did not change (Fig. 3.12).

Fig. 3.12. An enlargement of isopycnal contours for the second two days of the cruise, with PR mRNA concentration in the seawater overlaid. The two isopycnals represent σ_θ=23.812 ± 0.002 kg.m^{-3}, and σ_θ=23.824 ± 0.002 kg.m^{-3}, respectively.

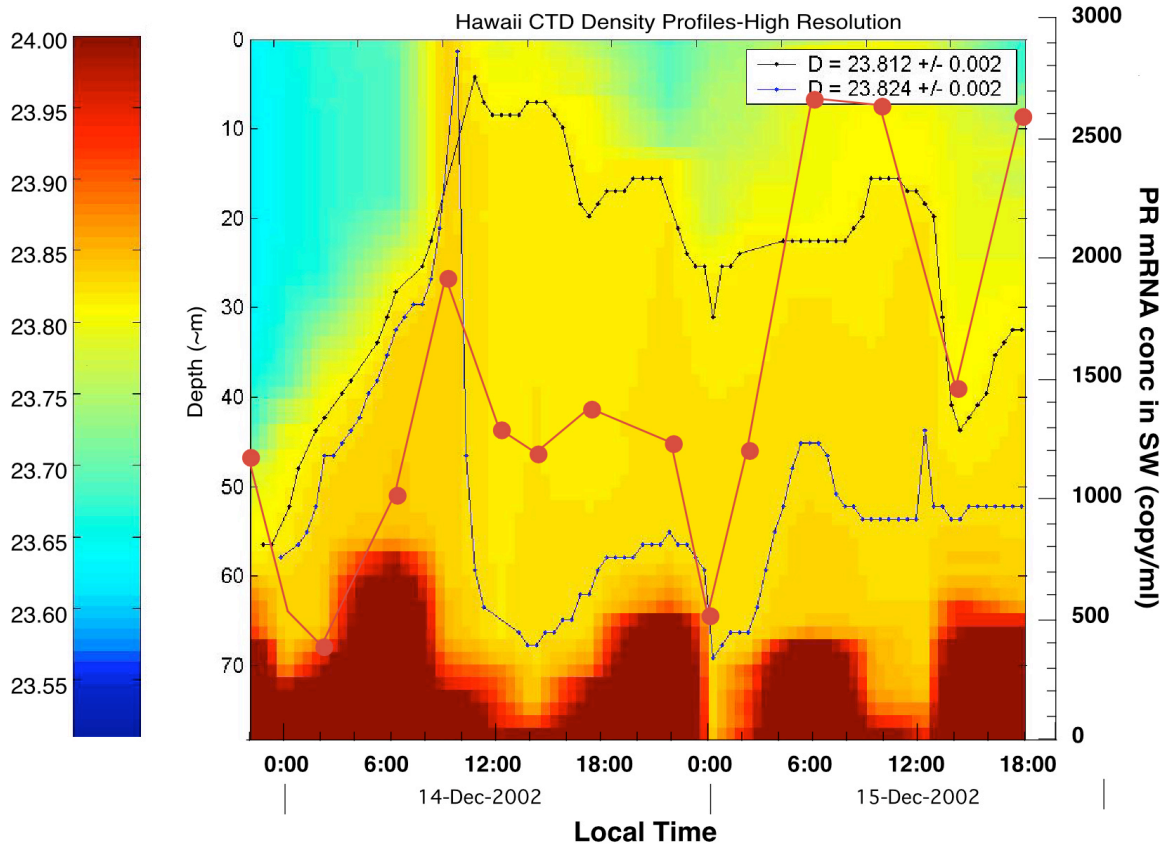
Discussion

Recent environmental genomics studies have provided an enormous amount of sequence information. Among these sequences, many novel genes have been

identified and associated with specific microbial activities (Béjà et al. 2000a; Schloss and Handelsman 2003). Even though regulatory mechanisms of these novel genes are mostly unknown, measurements of *in situ* gene transcription related to the environmental and biogeochemical parameters might provide insights into regulation and expression of these novel genes in the environment. In this study, we successfully developed a method for measuring microbial gene transcriptional activities in environmental samples using TaqMan assays, and used this method to study the expression regulation and dynamics of a group of PR genes in the oligotrophic North Pacific subtropical gyre.

As described above, several factors other than actual transcriptional activity might induce mRNA level changes, especially in highly complex and heterogeneous environmental samples, and these changes can be erroneously attributed to transcriptional level change. Our results showed that the normalization of target mRNA levels to target gene abundance can compensate for changes in cell abundance, since the increased concentrations of PR genes and mRNAs in the seawater (Fig. 3.3) did not lead to an increase of measured PR gene expression levels (Fig. 3.4A). Similar strategies were used to measure catechol-2,3-dioxygenase (*xyl*) gene expression in aquatic environments (Pichard and Paul 1993) and guanylate kinase (*gmk*) gene expression in Staphylococci (Vandecasteele et al. 2002). Furthermore, the assumption that the efficiency of nucleic extraction and reverse transcription were constant was likely true, since the RNA and DNA samples used in this study were extracted from bacterioplankton from 4 l (RNA) and 500 ml (DNA) of seawater from the Subtropical Pacific. These volumes were shown to be within the

ranges where RNA and DNA were shown to be extracted with constant efficiencies, and sufficient yields (Chapter 1).

We also observed that the number of PR mRNA copies per nanogram of total RNA significantly increased with time throughout the cruise (Fig. 3.4B). Rather than an increase in PR expression activity of the targeted organisms, we believe that this increase was caused by changes in bacterial community structure during the cruise, since total RNA does not reflect exclusively the abundance of the target microorganisms, but is a reflection of all microbial cells in the sample. Furthermore, the increased PR gene copies per nanogram of total DNA (Fig. 3.4C) also supported the hypothesis of community structure shifts, and when used to normalize PR mRNA abundance per nanogram of total RNA, resulted in a similar diel profile of PR gene expression to that obtained using per gene normalization approach (Fig. 3.4D), further supporting the use of target gene as a normalizer in gene expression studies.

The isohaline contours (Fig. 3.6) reveal an obvious high salinity intrusion event after 12:00 AM on December 14, 2002 of the cruise, which might have contributed to changes in the bacterial community at the sampling site. Total *Bacteria* and SAR11 16S rRNA genes were quantified by real time PCR and the results further supported our hypothesis of community structure change. *Bacteria* 16S rRNA gene abundance showed a significant increase after the second day of the cruise (Fig. 3.8A) (Student's t test, $p=0.003$). Furthermore, since the targeted PR genes are possibly harbored by a bacterium belonging to the SAR11 clade (Chapter 2), the significant increase of the ratio of targeted PR genes to SAR11 rRNA genes indicates possible changes within the SAR11 clade itself.

Transcriptional level regulation of PR genes by light has been previously suggested by the presence of regulatory sites in the sequences of the originally described PR-containing BAC clone (Baliga et al. 2001), as well as in expression studies of Green-absorbing PR in the Monterey Bay (Suzuki et al., unpublished). The observation that the PR expression peaked in early mornings (likely because light condition is favorable for PR synthesis), decreased towards late afternoons (likely because synthesized PRs are sufficient for cell function with the high irradiance), and reached minima at midnights (likely because no light is available to induce PR synthesis), indicates possible light effects on PR transcription and gene regulation at transcriptional level. However, late night (~10 pm) peaks were also observed in two successive days, which could not be explained by light-driven transcriptional activity. We found that these late night peaks of mRNA-to-gene ratio were more likely due to the steep decrease of PR gene abundance rather than the increase of PR transcript (Fig. 3.3). The fact that this decreasing PR gene concentration was associated with an apparent upward water intrusion (Fig. 3.10) indicates possible effects of physical processes on the measurements of PR gene abundance, PR mRNA abundance and thus PR gene expression levels.

The results of a more in depth investigation of hydrographic conditions at the sampling site further supported the influence of physical forcing throughout the cruise. Internal tide signals could be observed at deeper waters (>100 m) in the salinity field (Fig. 3.6), but the higher salinity water intrusion complicated and masked the internal tide signal at shallower depths. Fortunately, a higher resolution analysis of the density field (Fig. 3.8A) allowed visualization of physical processes

around the 25 m depth after the intrusion, when isopycnal compressions as well as internal tide signals were observed. Increased fluorescence values have been observed on the eddy edges on the Central California Shelf, where shear and isopycnal compression caused thinning in the vertical dimension and compression of biomass during the formation of thin layers (Margaret A. McManus, University of Hawaii, personal communication). Similarly, our results also indicated higher total nucleic acid concentrations and higher PR gene concentrations during isopycnal compressions (Fig. 3.8). In contrast, three out of four isopycnal compressions were associated with PR mRNA concentration decrease rather than increase (Fig. 3.11), indicating decreased PR transcriptional activities at the three time points (12:31 PM, 12:25 AM, 2:31 PM). It is worth noting that these three time points are associated with considerably high (12:31 PM and 2:31 PM) or low (12:25 AM) irradiance that would cause less active PR transcriptional activity were PR expression regulated by light, supporting the effects of light in PR expression.

Interestingly, the steep decrease of the PR gene concentration at 10 PM on December 14, 2002 that was associated with the 10 PM PR expression peak, occurred shortly prior to an occurrence of isopycnal compression, and was followed by a sharp increase of PR gene concentration matching the isopycnal compression (Fig. 3.9). This remarkable rapid fluctuation was attributed to internal tide effects revealed by a quick shallowing of isopycnal surfaces (Fig. 3.10). Analogous to the well-known Deep Chlorophyll Maximum (DCM), we hypothesize that in upper ocean, PR gene concentration (an indicator of cell numbers) decreases with depth while PR mRNA concentration increases with depth, in response to light limitation. Therefore,

shallowing of isopycnal surfaces are associated with lower PR gene and higher PR mRNA concentrations, and deepening of isopycnal surfaces are associated with higher PR gene and lower PR mRNA concentrations. This effect that also indicates possible light effects on PR transcription, and gene regulation at transcriptional level was partly supported by four depth profiles of PR gene expression (per gene normalization) where increases in PR mRNAs and decreases in PR genes from 25 m to 40 m was observed in all profiles, although the same trend was not observed from the surface to 25 m (data not shown).

Following the higher salinity intrusion, PR gene concentrations responded to isopycnal fluctuations as predicted by the proposed model, while two exceptions were observed for PR mRNA concentrations. Both exceptions occurred when the PR mRNA concentrations were expected to decrease with the isopycnal deepening but instead, mRNA concentrations increased or maintained the same levels. Interestingly, both samples with the exceptional mRNA concentrations were collected at times (5:30 PM and 10:10 AM) when PR gene expression would be expected to be higher were it to be regulated by light.

In conclusion, the measurement of *in situ* expression for genes harbored by not-yet-cultured microbes is highly challenging. If not properly normalized, changes in abundance of target gene-containing organisms and in microbial community structure may cause biased gene expression profiles. Furthermore, various environmental factors are likely to affect the measured values, some of which may control expression changes in coherent microbial populations, while others, like advection, might cause non expression-driven changes of measured gene transcription

and lead to data misinterpretation, and thus should be carefully accounted for. Although cultivation is ultimately necessary for a complete and comprehensive understanding of gene expression mechanisms in living cells, the per gene normalization approach developed herein allows biologically relevant measurement of *in situ* transcriptional activities of genes in the environment and, when related to environmental variables, provides novel insights into further studies on microbial gene regulation, where no cultured models are available.

Literature cited

Baliga NS, Kennedy SP, Ng WV, Hood L and DasSarma S (2001). "Genomic and genetic dissection of an archaeal regulon." <u>Proc Natl Acad Sci U S A</u> 98(5): 2521-5.

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN and DeLong EF (2000a). "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea." <u>Science</u> 289(5486): 1902-6.

Bustin SA (2002). "Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems." <u>J Mol Endocrinol</u> 29(1): 23-39.

Corredor JE, Wawrik B, Paul JH, Tran H, Kerkhof L, Lopez JM, Dieppa A and Cardenas O (2004). "Geochemical rate-RNA integrated study: Ribulose-1,5-bisphosphate carboxylase/oxygenase gene transcription and photosynthetic capacity of planktonic photoautotrophs." <u>Appl Environ Microbiol</u> 70(9): 5459-5468.

Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G and Zumla A (2004). "Validation of housekeeping genes for normalizing RNA expression in real-time PCR." <u>Biotechniques</u> 37(1): 112-4, 116, 118-9.

Gibson UE, Heid CA and Williams PM (1996). "A novel method for real time quantitative RT-PCR." <u>Genome Res</u> 6(10): 995-1001.

Giovannoni SJ and Rappé MS (2000). Evolution, Diversity, and Molecular Ecology of Marine Prokaryotes. <u>Microbial Ecology of the Oceans</u>. Kirchman, DL. New

York, Wiley-Liss: 47-84.

Holland PM, Abramson RD, Watson R and Gelfand DH (1991). "Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase." Proc Natl Acad Sci U S A 88(16): 7276-80.

Kaeberlein T, Lewis K and Epstein SS (2002). "Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment." Science 296(5570): 1127-1129.

Leclerc GJ, Leclerc GM and Barredo JC (2002). "Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines." Cancer Cell Int 2(1): 1.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A and Schleifer KH (2004). "ARB: a software environment for sequence data." Nucleic Acids Res 32(4): 1363-71.

Miller TJ, Herra T and Leggett WC (1995). "An Individual-Based Analysis of the Variability of Eggs and Their Newly-Hatched Larvae of Atlantic Cod (Gadus-Morhua) on the Scotian Shelf." Canadian Journal of Fisheries and Aquatic Sciences 52(5): 1083-1093.

Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR and Stahl DA (1986). "Microbial ecology and evolution: a ribosomal RNA approach." Annu Rev Microbiol 40: 337-65.

Pace NR (1997). "A molecular view of microbial diversity and the biosphere." Science 276(5313): 734-40.

Parker RM and Barnes NM (1999). "mRNA: detection by in Situ and northern hybridization." Methods Mol Biol 106: 247-83.

Pichard SL and Paul JH (1993). "Gene-Expression Per Gene Dose, a Specific Measure of Gene-Expression in Aquatic Microorganisms." Appl Environ Microbiol 59(2): 451-457.

Rappé MS, Connon SA, Vergin KL and Giovannoni SJ (2002). "Cultivation of the ubiquitous SAR11 marine bacterioplankton clade." Nature 418(6898): 630-3.

Schloss PD and Handelsman J (2003). "Biotechnological prospects from metagenomics." Curr Opin Biotechnol 14(3): 303-10.

Shand RF and Betlach MC (1991). "Expression of the bop gene cluster of Halobacterium halobium is induced by low oxygen tension and by light." J Bacteriol 173(15): 4692-9.

Sharkey FH, Markos SI and Haylock RW (2004b). "Quantification of toxin-encoding mRNA from Clostridium botulinum type E in media containing sorbic acid or sodium nitrite by competitive RT-PCR." FEMS Microbiol Lett 232(2): 139-44.

Ståhlberg A, Kubista M and Pfaffl M (2004). "Comparison of reverse transcriptases in gene expression analysis." Clin Chem 50(9): 1678-1680.

Stein JL, Marsh TL, Wu KY, Shizuya H and DeLong EF (1996). "Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment front a planktonic marine archaeon." J Bacteriol 178(3): 591-599.

Suzuki MT and Giovannoni SJ (1996). "Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR." Appl Environ Microbiol 62(2): 625-30.

Suzuki MT, Taylor LT and DeLong EF (2000). "Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays." Appl Environ Microbiol 66(11): 4605-14.

Suzuki MT, Preston CM, Chavez FP and DeLong EF (2001b). "Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay." Aquatic Microbial Ecology 24(2): 117-127.

Tomita H, Kotake Y and Anderson RE (2005). "Mechanism of protection from light-induced retinal degeneration by the synthetic antioxidant phenyl-n-tert-butylnitrone." Investigative Ophthalmology & Visual Science 46(2): 427-434.

Vandecasteele SJ, Peetermans WE, Merckx R, Van Ranst M and Van Eldere J (2002). "Use of gDNA as internal standard for gene expression in staphylococci in vitro and in vivo." Biochem Biophys Res Commun 291(3): 528-34.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H,

Pfannkoch C, Rogers YH and Smith HO (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." <u>Science</u> 304(5667): 66-74.

Walker NJ (2002). "A technique whose time has come." <u>Science</u> 296(5567): 557-559.

Wawrik B, Paul JH and Tabita FR (2002). "Real-time PCR quantification of rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes." <u>Appl Environ Microbiol</u> 68(8): 3771-9.

Weinbauer MG, Fritz I, Wenderoth DF and Hofle MG (2002). "Simultaneous extraction from bacterioplankton of total RNA and DNA suitable for quantitative structure and function analyses." <u>Appl Environ Microbiol</u> 68(3): 1082-7.

Wilson MS, Bakermans C and Madsen EL (1999). "In situ, real-time catabolic gene expression: Extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater." <u>Appl Environ Microbiol</u> 65(1): 80-87.

Wyman M, Zehr JP and Capone DG (1996). "Temporal variability in nitrogenase gene expression in natural populations of the marine cyanobacterium Trichodesmium thiebautii." <u>Appl Environ Microbiol</u> 62(3): 1073-1075.

Wyman M (1999). "Diel rhythms in ribulose-1,5-bisphosphate carboxylase/oxygenase and glutamine synthetase gene expression in a natural population of marine picoplanktonic cyanobacteria (Synechococcus spp.)." <u>Appl Environ Microbiol</u> 65(8): 3651-9.

Wyman M, Davies JT, Hodgson S, Tarran GA and Purdie DA (2005). "Dynamics of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase Gene Expression in the Coccolithophorid Coccolithus pelagicus during a Tracer Release Experiment in the Northeast Atlantic." <u>Appl Environ Microbiol</u> 71(3): 1659-61.

Zengler K, Toledo G, Rappé MS, Elkins J, Mathur EJ, Short JM and Keller M (2002). "Cultivating the uncultured." <u>Proc Natl Acad Sci U S A</u> 99(24): 15681-6.

Zhong H and Simons JW (1999). "Direct comparison of GAPDH, beta-actin, cyclophilin, and 28S rRNA as internal standards for quantifying RNA levels under hypoxia." <u>Biochem Biophys Res Commun</u> 259(3): 523-6.

Conclusions and Recommendations

Here, I summarize the results obtained from three studies that compose this thesis: the investigation of RNA extraction efficiency constancy; the phylogenetic analysis of actively expressed PR genes in the North Pacific Central Gyre; and the measurement of *in situ* PR expression in a three-day diel time series, described in details in Chapter 1, 2, and 3, respectively. I also discuss the significance of the RNA extraction method and microbial gene expression measurement approach developed and optimized in this study to the understanding of specific microbial activities and controlling factors in biogeochemical processes of marine ecosystems.

Constancy of RNA extraction efficiency

The results of the experiment evaluating constancy of total RNA and specific mRNA (PR mRNA) extraction efficiency suggested that quasi-absolute quantification of microbial genes and mRNAs in environmental samples is feasible. Microbial ecologists have long recognized that quantifying specific genes (or mRNAs) relative to community-wide parameters such as total nucleic acids or rRNA gene abundance faces the risk of data misinterpretation due to possible microbial community change. Apparent increases (or decreases) of target gene (mRNA) signal might just be resulting from decreases (or increases) of non-target gene (or mRNA and rRNA) abundance. An alternative solution to this problem is the absolute quantification of target gene in the environment (i.e., measurement of gene (or mRNA) copy numbers per volume of seawater).

The optimized RNA extraction method developed here enables constant RNA extraction efficiencies for bacterioplankton samples filtered from predetermined volumes of water. This volume range is dependent on sample type, as indicated by

our results from the North Pacific and Patuxent Estuary samples. Nonetheless, we find that bacterioplankton samples containing specific amounts of RNA, regardless of sample types, can be extracted at a relatively constant efficiency using the optimized RNA extraction method. Therefore, it is important to evaluate a volume range from which bacterioplankton samples should be filtered in order to gain constant extraction efficiency, based on cell abundance. Although the evaluation of specific mRNA extraction efficiency was complicated by the uncertainty of Reverse Transcription (RT) efficiency, we were able to find a linear relationship between quantified PR mRNA copy number and filtered water volume for some of tested subsamples, indicating that, for these subsamples, both mRNA extraction and RT efficiencies were relatively constant. However, the optimization of RT reactions and the improvement of the range of constant RT efficiency remain as subjects for future research.

In conclusion, quasi-absolute quantification of microbial genes and mRNAs is possible, but optimized DNA and RNA extraction methods that allow constant extraction efficiency, and optimized RT reactions are required. These issues are of great relevance for the measurement of expression of functional genes harbored by not-yet-cultured microbes in the environment, since quasi-absolute quantification of specific genes and mRNAs validates the use of mRNA-over-gene ratio as a measure of gene expression levels.

Phylogenetic analysis of actively expressed PR genes

Previous studies have indicated a high diversity of proteorhodopsin (PR) genes in the environment, via PCR-based gene surveys, large insert clone library

screening, and environmental shotgun sequencing efforts (Béjà et al. 2000a; de la Torre et al. 2003; Sabehi et al. 2004; Venter et al. 2004). Particularly, since different PR genes have been retrieved from different marine ecosystems, including the North and Eastern Pacific Ocean, the Atlantic Ocean, the Southern Ocean, and the Mediterranean, Red and Sargasso Seas, it is very likely that PR-expressing organisms predominant in different marine ecosystems are not the same.

To our knowledge, this was the first effort exploring *in situ* PR gene expression based on PR mRNA diversity and abundance in the North Pacific Central Gyre, and thus no prior knowledge of the identity and diversity of actively expressed PR genes was available. The sequencing and analysis of the cDNA clone library constructed with degenerate primers targeting diverse PR genes indicated that diverse bacteria harbored and actively expressed PR genes at 25 m depth at Station ALOHA. PR sequences associated with SAR11 or SAR11-like bacteria in particular were prevalent in the cDNA clone library, indicating that these organisms might represent important PR-expressing organisms in the photic zone of the Central North Pacific.

The main problems with the reverse transcription of PR mRNA were associated with the low abundance of target mRNA in the highly complex RNA mixtures extracted from environmental bacterioplankton samples. Reverse transcription reactions using random primers failed to produce detectable PCR products, and RT reactions using target specific degenerate primers at a lower temperature resulted in non-specific PCR amplification. Only RT reactions using degenerate primers and thermo-stable reverse transcriptase at an elevated temperature led to the specific amplification of PR cDNAs. Furthermore, primer driven biases in

RT and PCR procedures are still possible but not likely to be severe, since the highly degenerate primers were also used to retrieve different PR sequences in other marine ecosystems (Sabehi et al. 2005).

It was operationally difficult to design real-time PCR assays (TaqMan in this study) for all PR cDNA types retrieved in the clone library since they were highly diverse. Therefore, a PR group that was tightly clustered and closely related to a PR gene harbored by a SAR11 strain was chosen as the target of TaqMan assays in the subsequent gene expression experiment.

Measurement of *in situ* gene expression in the environment

The results of the measurement of *in situ* PR transcript levels suggest possible light regulation on PR gene transcriptional activities, further indicating that the expression of PR genes might be regulated, at least in part, at the transcriptional level, and thus PR mRNA levels can be used as surrogates of PR expression levels. In this study, we normalized PR mRNA abundance to PR gene abundance in order to avoid biases due to possible changes in abundance of the targeted organisms. Unbiased quantification of PR mRNAs and PR genes was a critical factor to retain the biological relevance of this novel approach of measuring *in situ* gene expression. As discussed in the RNA extraction section above, PR mRNAs and genes were quantified in a "quasi-absolute" fashion as copies per volume seawater, assuming that the extraction efficiency and reverse transcription efficiency were 100%.

I am aware that cell cycle-related activities such as cell proliferation and possible bacterial synchronous growth in the environment might have affected measured gene expression levels, by the duplication of target genes and the resulting

underestimation of gene expression level. Although synchronous growth of *Synechococcus* and *Prochlorococcus* populations in natural waters have been suggested in previous studies (Holtzendorff et al. 2001; Vaulot et al. 1996), to date we are only aware of the existence of one isolated organism containing PR, and we were unable to find evidence for synchronous growth of PR-containing bacteria in the literature, and thus this remains as a subject for future investigations.

Environmental controlling factors of measured gene expression levels

The results of the study comparing diel dynamics of measured PR gene expression levels and environmental parameters (light, salinity, temperature and density) indicated light effects on PR expression, and suggested that physical processes like advections affected the measured levels of PR gene expression.

Measuring *in situ* microbial gene expression and the subsequent data interpretation was challenging, since some environmental factors regulate the actual expression level of target genes, while others contribute to the variation of measured gene expression, leading to possible data misinterpretation. In the case of PR genes, light was hypothesized as a controlling factor of PR gene expression, since PRs absorb light and function as proton pumps to produce cellular energy. However, in addition to early morning samples, 10 P.M. samples showed increased PR expression levels, which could not be explained solely by regulation due to light availability. Rather, it was suggested to be a response of decreased PR gene abundance associated with upward water movement advecting less abundant PR-harboring bacteria with higher PR mRNA levels to 25 m, indicating the effects of other physical processes on

the measured PR gene expression. It is important to point out that, the PR mRNA to gene ratios of the different isopycnals crossing the 25 m isobath are in accordance to light effects on PR expression.

In conclusion, this study was the first effort, to our knowledge, made to identify and measure the *in situ* expression levels of actively expressed PR genes in the North Pacific Central Gyre. Using the novel approach proposed in this study, we were able to obtain biologically relevant measurements for *in situ* PR gene expression, which, when combined with environmental parameters such as light and hydrographic conditions, provided novel insights into possible mechanisms of PR gene expression regulation. Finally, the strategies employed in this study, including nucleic acid extraction with constant efficiency, cDNA clone library construction from environmental samples, and per gene normalization for target transcripts, circumvent specific problems commonly present in cultivation-independent microbial studies and possess application potentials for such studies in the future.

Recommendations

In order to further confirm our hypotheses that oceanic physical processes affected measured gene expression levels, satellite picture of the hydrographic conditions in the region and light profiles during the cruise should be combined in future work. A fine resolution depth profile would also be helpful. Although four depth profiles were collected together with the diel samples, they were collected at depths that were at least 20 meters apart, hindering the detailed investigation of PR expression with depth.

Bibliography

Azam F (1998). "Microbial control of oceanic carbon flux: The plot thickens." Science **280**(5364): 694-696.

Azam F and Worden AZ (2004). "Oceanography. Microbes, molecules, and marine ecosystems." Science **303**(5664): 1622-4.

Baliga NS, Kennedy SP, Ng WV, Hood L and DasSarma S (2001). "Genomic and genetic dissection of an archaeal regulon." Proc Natl Acad Sci U S A **98**(5): 2521-5.

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN and DeLong EF (2000a). "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea." Science **289**(5486): 1902-6.

Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, Villacorta R, Amjadi M, Garrigues C, Jovanovich SB, Feldman RA and DeLong EF (2000b). "Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage." Environ Microbiol **2**(5): 516-29.

Béjà O, Spudich EN, Spudich JL, Leclerc M and DeLong EF (2001). "Proteorhodopsin phototrophy in the ocean." Nature **411**(6839): 786-9.

Béjà O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T, Eisen JA, Fraser CM and DeLong EF (2002). "Unsuspected diversity among marine aerobic anoxygenic phototrophs." Nature **415**(6872): 630-3.

Bostrom KH, Simu K, Hagstrom A and Riemann L (2004). "Optimization of DNA extraction for quantitative marine bacterioplankton community analysis." Limnology and Oceanography-Methods **2**: 365-373.

Bustin SA (2000). "Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays." J Mol Endocrinol **25**(2): 169-93.

Bustin SA (2002). "Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems." J Mol Endocrinol **29**(1): 23-39.

Corredor JE, Wawrik B, Paul JH, Tran H, Kerkhof L, Lopez JM, Dieppa A and Cardenas O (2004). "Geochemical rate-RNA integrated study: Ribulose-1,5-bisphosphate carboxylase/oxygenase gene transcription and photosynthetic capacity of planktonic photoautotrophs." Appl Environ Microbiol **70**(9): 5459-5468.

Curry J, McHale C and Smith MT (2002). "Low Efficiency of the Moloney Murine Leukemia Virus Reverse Transcriptase during Reverse Transcription of Rare t(8;21) Fusion Gene Transcripts." Biotechniques **32**(4): 768-775.

de la Torre JR, Christianson LM, Béjà O, Suzuki MT, Karl DM, Heidelberg J and DeLong EF (2003). "Proteorhodopsin genes are distributed among divergent marine bacterial taxa." Proc Natl Acad Sci U S A **100**(22): 12830-5.

DeLong EF (2004). "Microbial population genomics and ecology: the road ahead." Environ Microbiol **6**(9): 875-8.

Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G and Zumla A (2004). "Validation of housekeeping genes for normalizing RNA expression in real-time PCR." Biotechniques **37**(1): 112-4, 116, 118-9.

Doney SC, Abbott MR, Cullen JJ, Karl DM and Rothstein L (2004). "From genes to ecosystems: the ocean's new frontier." Front Ecol Environ **2**(9): 457-466.

Eichler S, Weinbauer MG, Dominik K and Hofle MG (2004). Extraction of total RNA and DNA from bacterioplankton. Molecular Microbial Ecology Manual. Kowalchuk, GA, de Bruijn, FJ, Head, IM, Akkermans, AD and van Elsas, JD. Dordrecht, the Netherlands, Kluwer Acad. Publ.**:** 1780.

Fey A, Eichler S, Flavier S, Christen R, Hofle MG and Guzman CA (2004). "Establishment of a real-time PCR-based approach for accurate quantification of bacterial RNA targets in water, using Salmonella as a model organism." Appl Environ Microbiol **70**(6): 3618-23.

Gibson UE, Heid CA and Williams PM (1996). "A novel method for real time quantitative RT-PCR." Genome Res **6**(10): 995-1001.

Giovannoni SJ and Rappé MS (2000). Evolution, Diversity, and Molecular Ecology of Marine Prokaryotes. Microbial Ecology of the Oceans. Kirchman, DL. New York, Wiley-Liss**:** 47-84.

Golding GB (1983). "Estimates of DNA and protein sequence divergence: an examination of some assumptions." Mol Biol Evol **1**(1): 125-42.

Holland PM, Abramson RD, Watson R and Gelfand DH (1991). "Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase." Proc Natl Acad Sci U S A **88**(16): 7276-80.

Holtzendorff J, Partensky F, Jacquet S, Bruyant F, Marie D, Garczarek L, Mary I, Vaulot D and Hess WR (2001). "Diel expression of cell cycle-related genes in synchronized cultures of Prochlorococcus sp. strain PCC 9511." J Bacteriol **183**(3): 915-20.

Huelsenbeck JP, Ronquist F, Nielsen R and Bollback JP (2001). "Bayesian inference of phylogeny and its impact on evolutionary biology." Science **294**(5550): 2310-4.

Jin L and Nei M (1990). "Limitations of the evolutionary parsimony method of phylogenetic analysis." Mol Biol Evol **7**(1): 82-102.

Jones DT, Taylor WR and Thornton JM (1992). "The rapid generation of mutation data matrices from protein sequences." Comput Appl Biosci **8**(3): 275-82.

Kaeberlein T, Lewis K and Epstein SS (2002). "Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment." Science **296**(5570): 1127-1129.

Leclerc GJ, Leclerc GM and Barredo JC (2002). "Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines." Cancer Cell Int **2**(1): 1.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A and Schleifer KH (2004). "ARB: a software environment for sequence data." Nucleic Acids Res **32**(4): 1363-71.

Man-Aharonovich D, Sabehi G, Sineshchekov OA, Spudich EN, Spudich JL and Béjà O (2004). "Characterization of RS29, a blue-green proteorhodopsin variant from the Red Sea." Photochem Photobiol Sci **3**(5): 459-62.

Miller TJ, Herra T and Leggett WC (1995). "An Individual-Based Analysis of the Variability of Eggs and Their Newly-Hatched Larvae of Atlantic Cod (Gadus-Morhua) on the Scotian Shelf." Canadian Journal of Fisheries and Aquatic Sciences **52**(5): 1083-1093.

Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA and Giovannoni SJ (2002). "SAR11 clade dominates ocean surface bacterioplankton communities." Nature **420**(6917): 806-10.

Notredame C, Higgins DG and Heringa J (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." J Mol Biol **302**(1): 205-17.

Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR and Stahl DA (1986). "Microbial ecology and evolution: a ribosomal RNA approach." Annu Rev Microbiol **40**: 337-65.

Pace NR (1997). "A molecular view of microbial diversity and the biosphere." Science **276**(5313): 734-40.

Parker RM and Barnes NM (1999). "mRNA: detection by in Situ and northern hybridization." Methods Mol Biol **106**: 247-83.

Paul JH and Pichard SL (1995). Extraction of DNA and RNA from Aquatic Environments. Nucleic Acids in the Environment. Trevors, JT and van Elsas, JD. New York, Springer**:** 153-177.

Pichard SL and Paul JH (1991). "Detection of Gene-Expression in Genetically Engineered Microorganisms and Natural Phytoplankton Populations in the Marine-Environment by Messenger-Rna Analysis." Appl Environ Microbiol **57**(6): 1721-1727.

Pichard SL and Paul JH (1993). "Gene-Expression Per Gene Dose, a Specific Measure of Gene-Expression in Aquatic Microorganisms." Appl Environ Microbiol **59**(2): 451-457.

Rappé MS, Connon SA, Vergin KL and Giovannoni SJ (2002). "Cultivation of the ubiquitous SAR11 marine bacterioplankton clade." Nature **418**(6898): 630-3.

Rodriguez-Valera F (2004). "Environmental genomics, the big picture?" FEMS Microbiol Lett **231**(2): 153-8.

Ronquist F and Huelsenbeck JP (2003). "MrBayes 3: Bayesian phylogenetic inference under mixed models." Bioinformatics **19**(12): 1572-4.

Sabehi G, Massana R, Bielawski JP, Rosenberg M, Delong EF and Béjà O (2003). "Novel Proteorhodopsin variants from the Mediterranean and Red Seas." Environ Microbiol **5**(10): 842-9.

Sabehi G, Béjà O, Suzuki MT, Preston CM and DeLong EF (2004). "Different SAR86 subgroups harbour divergent proteorhodopsins." <u>Environ Microbiol</u> **6**(9): 903-10.

Sabehi G, Loy A, Jung KH, Partha R, Spudich JL, Isaacson T, Hirschberg J, Wagner M and Béjà O (2005). "New Insights into Metabolic Properties of Marine Bacteria Encoding Proteorhodopsins." <u>PLoS Biol</u> **3**(8): e273.

Schloss PD and Handelsman J (2003). "Biotechnological prospects from metagenomics." <u>Curr Opin Biotechnol</u> **14**(3): 303-10.

Shand RF and Betlach MC (1991). "Expression of the bop gene cluster of Halobacterium halobium is induced by low oxygen tension and by light." <u>J Bacteriol</u> **173**(15): 4692-9.

Sharkey FH, Banat IM and Marchant R (2004a). "Detection and quantification of gene expression in environmental bacteriology." <u>Appl Environ Microbiol</u> **70**(7): 3795-806.

Sharkey FH, Markos SI and Haylock RW (2004b). "Quantification of toxin-encoding mRNA from Clostridium botulinum type E in media containing sorbic acid or sodium nitrite by competitive RT-PCR." <u>FEMS Microbiol Lett</u> **232**(2): 139-44.

Sohail M and Southern EM (2000). "Hybridization of antisense reagents to RNA." <u>Curr Opin Mol Ther</u> **2**(3): 264-71.

Ståhlberg A, Hakansson J, Xian X, Semb H and Kubista M (2004a). "Properties of the reverse transcription reaction in mRNA quantification." <u>Clin Chem</u> **50**(3): 509-15.

Ståhlberg A, Kubista M and Pfaffl M (2004b). "Comparison of reverse transcriptases in gene expression analysis." <u>Clin Chem</u> **50**(9): 1678-1680.

Staley JT and Konopka A (1985). "Measurement of *in situ* Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats." <u>Annual Review of Microbiology</u> **39**: 321-346.

Stapels MD, Cho JC, Giovannoni SJ and Barofsky DF (2004). "Proteomic analysis of novel marine bacteria using MALDI and ESI mass spectrometry." <u>J Biomol Tech</u> **15**(3): 191-8.

Stein JL, Marsh TL, Wu KY, Shizuya H and DeLong EF (1996). "Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment front a planktonic marine archaeon." <u>J Bacteriol</u> **178**(3): 591-599.

Suzuki MT and Giovannoni SJ (1996). "Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR." <u>Appl Environ Microbiol</u> **62**(2): 625-30.

Suzuki MT, Taylor LT and DeLong EF (2000). "Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays." <u>Appl Environ Microbiol</u> **66**(11): 4605-14.

Suzuki MT, Béjà O, Taylor LT and Delong EF (2001a). "Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton." <u>Environ Microbiol</u> **3**(5): 323-31.

Suzuki MT, Preston CM, Chavez FP and DeLong EF (2001b). "Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay." <u>Aquatic Microbial Ecology</u> **24**(2): 117-127.

Tomita H, Kotake Y and Anderson RE (2005). "Mechanism of protection from light-induced retinal degeneration by the synthetic antioxidant phenyl-n-tert-butylnitrone." <u>Investigative Ophthalmology & Visual Science</u> **46**(2): 427-434.

Torsvik V, Ovreas L and Thingstad TF (2002). "Prokaryotic diversity--magnitude, dynamics, and controlling factors." <u>Science</u> **296**(5570): 1064-6.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS and Banfield JF (2004). "Community structure and metabolism through reconstruction of microbial genomes from the environment." <u>Nature</u> **428**(6978): 37-43.

Vandecasteele SJ, Peetermans WE, Merckx R and Van Eldere J (2001). "Quantification of expression of Staphylococcus epidermidis housekeeping genes with TaqMan quantitative PCR during in vitro growth and under different conditions." <u>J Bacteriol</u> **183**(24): 7094-101.

Vandecasteele SJ, Peetermans WE, Merckx R, Van Ranst M and Van Eldere J (2002). "Use of gDNA as internal standard for gene expression in staphylococci in vitro and in vivo." <u>Biochem Biophys Res Commun</u> **291**(3): 528-34.

Vaulot D, LeBot N, Marie D and Fukai E (1996). "Effect of phosphorus on the Synechococcus cell cycle in surface Mediterranean waters during summer." Appl Environ Microbiol **62**(7): 2527-2533.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH and Smith HO (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." Science **304**(5667): 66-74.

Walker NJ (2002). "A technique whose time has come." Science **296**(5567): 557-559.

Ward DM, Ruff-Roberts AL and Weller R (1995). Methods for extracting RNA or ribosomes from microbial mats and cultivated microorganisms. Molecular Microbial Ecology Manual. Kowalchuk, GA, de Bruijn, FJ, Head, IM, Akkermans, AD and van Elsas, JD. Dordrecht, the Netherlands, Kluwer Acad. Publ.**:** 1780.

Wawer C and Muyzer G (1995). "Genetic diversity of Desulfovibrio spp. in environmental samples analyzed by denaturing gradient gel electrophoresis of [NiFe] hydrogenase gene fragments." Appl Environ Microbiol **61**(6): 2203-10.

Wawrik B, Paul JH and Tabita FR (2002). "Real-time PCR quantification of rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes." Appl Environ Microbiol **68**(8): 3771-9.

Weinbauer MG, Fritz I, Wenderoth DF and Hofle MG (2002). "Simultaneous extraction from bacterioplankton of total RNA and DNA suitable for quantitative structure and function analyses." Appl Environ Microbiol **68**(3): 1082-7.

Whelan S and Goldman N (2001). "A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach." Mol Biol Evol **18**(5): 691-9.

Whelan S, Lio P and Goldman N (2001). "Molecular phylogenetics: state-of-the-art methods for looking into the past." Trends Genet **17**(5): 262-72.

Wilson MS, Bakermans C and Madsen EL (1999). "In situ, real-time catabolic gene expression: Extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater." Appl Environ Microbiol **65**(1): 80-87.

Wyman M, Zehr JP and Capone DG (1996). "Temporal variability in nitrogenase gene expression in natural populations of the marine cyanobacterium Trichodesmium thiebautii." <u>Appl Environ Microbiol</u> **62**(3): 1073-1075.

Wyman M (1999). "Diel rhythms in ribulose-1,5-bisphosphate carboxylase/oxygenase and glutamine synthetase gene expression in a natural population of marine picoplanktonic cyanobacteria (Synechococcus spp.)." <u>Appl Environ Microbiol</u> **65**(8): 3651-9.

Wyman M, Davies JT, Hodgson S, Tarran GA and Purdie DA (2005). "Dynamics of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase Gene Expression in the Coccolithophorid Coccolithus pelagicus during a Tracer Release Experiment in the Northeast Atlantic." <u>Appl Environ Microbiol</u> **71**(3): 1659-61.

Yang ZH (1994). "Maximum-Likelihood Phylogenetic Estimation from DNA-Sequences with Variable Rates over Sites - Approximate Methods." <u>Journal of Molecular Evolution</u> **39**(3): 306-314.

Yang ZH (1996). "Among-site rate variation and its impact on phylogenetic analyses." <u>Trends in Ecology & Evolution</u> **11**(9): 367-372.

Zengler K, Toledo G, Rappé MS, Elkins J, Mathur EJ, Short JM and Keller M (2002). "Cultivating the uncultured." <u>Proc Natl Acad Sci U S A</u> **99**(24): 15681-6.

Zhong H and Simons JW (1999). "Direct comparison of GAPDH, beta-actin, cyclophilin, and 28S rRNA as internal standards for quantifying RNA levels under hypoxia." <u>Biochem Biophys Res Commun</u> **259**(3): 523-6.