7-13-2007

The Use of Microarrays in the Detection of the Gene Expression of Ribulose- 1,5- Bisphosphate Carboxylase/Oxygenase (RubisCO) in the Marine Environment

Kathryn Lafaye Bailey
University of South Florida

The Use of Microarrays in the Detection of the Gene Expression of Ribulose- 1,5- Bisphosphate Carboxylase/Oxygenase (RubisCO) in the Marine Environment

by

Kathryn Lafaye Bailey

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Science
College of Marine Science
University of South Florida

Major Professor: John H. Paul, III, Ph.D.
Ashanti J. Pyrtle, Ph.D.
Frank Pyrtle, Ph.D.

Date of Approval:
July 13, 2007

Keywords: Phytoplankton, CBB Pathway, Mississippi River Plume, *rbcL*, pelagophyte, *Synechococcus*

Dedication

This thesis is dedicated to my mother who has supported me in everything I ever wanted to do in life. Thank you for always being there to give me words of encouragement when I needed them and for the many swift and firm kicks in the behind that kept me going when I felt like giving up. I love you and I live to make you proud of me. I would also like to dedicate this work to the late Mrs. Adrienne Turner.

Acknowledgements

First and foremost, I want to thank God for His many blessings. I would also like to thank Him for placing the following people in my life, for without them, the completion of this degree would not have been possible: my advisor, Dr. John H. Paul for allowing me to work with him these past 2 years and for his guidance toward the completion of my research and thesis; Dr. Ashanti Pyrtle for writing the Bridge to the Doctorate Proposal that has funded me for the past two years and also for founding MSPHDS which has enriched my graduate school experience; Mr. Bernard Batson for that email he sent to me on July 26, 2004 that changed my life forever; Amy Long, Jennifer Mobberley, and Michelle Del la Rosa for befriending me when nobody else would; Erica Casper and Dr. Stacey Patterson for always "keeping it real"; Dr. Lauren McDaniel for all of the motherly advice when I could not reach my mother; and last but not least, Dr. David John for his patience and friendship.

Table of Contents

List of Tables

List of Figures

The Use of Microarrays in the Detection of the Gene Expression of Ribulose- 1,5-bisphosphate carboxylase/oxygenase (RubisCO) in the Marine Environment

Kathryn L. Bailey

ABSTRACT

The Calvin-Benson-Bassham (CBB) pathway is the primary pathway for the entry of inorganic carbon in the biosphere. Autotrophic organisms use this cycle to ultimately convert CO_2 into carbohydrates using a key enzyme known as ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO). The gene that encodes for the large subunit of RubisCO is *rbcL* and detection of its expression can be used to determine the autotrophic organisms present in the environment. Recently, microarrays have been used to study functional gene expression from environmental samples such as those obtained from sediments and soil. The purpose of this thesis is to combine microarray technology and *rbcL* expression analysis to investigate phytoplankton populations in the Mississippi River Plume (MRP). Initially, a macroarray was constructed to determine its capabilities of quantifying gene expression in MRP. PCR amplicons were spotted onto a nylon membrane and labeled transcript RNA was hybridized to each array. Due to the large amount of cross hybridization that was observed, a microarray was used. Microarray analysis revealed large amounts of *Synechococcus*, pelagophyte and prymnesiophyte expression in the surface waters. Furthermore, there was no chlorophte or

Prochlorococcus expression observed in the surface waters. Subsurface microarray data showed high levels of pelagophytes and other Form ID organisms. A significant chlorophyte signal was also observed in the subsurface. This study provides a third level of specificity at which phylogenetic diversity has been sampled in the MRP. Although a limited number of samples were analyzed by microarrays, this technology shows promise and this study was viewed as a pilot for their application. The *rbcL* probes designed were based upon published sequences from 2003 and we now have a much greater understanding of the diversity of *rbcL*-containing phytoplanktonic phylotypes. Future studies should employ this knowledge for judicious probe selection.

Chapter 1

Introduction

Environmental gene expression studies provide a means of determining what biological and biochemical activities are being facilitated by microorganisms in the environment. Gene expression is controlled by cellular responses to changes the cell's environment (44) and can be broken down into two processes: transcription and translation. Transcription is a process in which an enzyme, known as RNA polymerase, makes messenger RNA (mRNA), or an RNA copy, of gene(s) on a DNA strand. In translation, the mRNA is read, or translated, into an amino acid sequence to form protein via the action of ribosomes. The expression, or lack thereof, of genes in the ocean is dependent upon many environmental factors or stresses. These can include but are not limited to nutrient availability, salinity, pH and temperature. Any modifications to these factors can inhibit the transcription of a gene or increase the amount of expression and thereby affect the production of the protein that it encodes.

The measurement of the expression of discreet genes in the ocean is in its infancy. In a study of nitrogenase activity in *Trichodesmium sp.*, researchers found that transcription of *nifHDK* was on a diel pattern (7). The criteria for an endogenous rhythm are the ability to persist under constant environmental conditions, the ability to maintain a

cyclic pattern in different temperatures and the cycle must occur in a diel period (7, 50). The results showed that transcript production in *Tichodesmium sp.* maintained a relative constant 24 h cycle in uniform environmental conditions and with slight temperature change (7).

In another study, mRNA expression of a high-affinity phosphate transporter gene (*TcPHO*) of the prasinophyte, *Tetraselmis chui* was shown to be correlated to nutrient availability (8). Chung et al. (2003) extracted the total RNA from a growth-dependent subtracted *T. chui* cDNA library made from cultures both rich and depleted of nitrate and phosphate. Real time polymerase chain reaction (PCR) analysis showed that *TcPHO* mRNA expression in the phosphorous-replete cultures increased during all 4 days of the experiment indicating expression of *TcPHO* linked to phosphorus availability (8). Cultures replete in nitrogen showed a similar growth pattern to that of the phosphorus-replete cultures, however, the *TcPHO* mRNA expression levels of the low-nitrate cultures remained low throughout the course of the experiment (8).

1.1 Primary Production in the Ocean

Primary production is the synthesis and storage of organic molecules during the growth and reproduction of photosynthetic and other autotrophic organisms. Most oceanic primary production occurs in the photic zone which is located in the upper 200 m of the water column (10). Primary production in the ocean accounts for nearly half of the earth's primary production, which ranges from 35 to 65 Gt of carbon annually (10) thus resulting in the ocean acting as a large sink for CO_2 (44). Phytoplankton located in the

photic zone use photosynthesis to store energy and release oxygen back into the atmosphere and surrounding waters (27). Marine phytoplankton are responsible for approximately half of Earth's annual oxygen production (44). During photosynthesis, CO_2 is taken up by phytoplankton and/or autotrophic bacteria and incorporated into their cells with oxygen and water as byproducts. The overall reaction for photosynthesis is $nCO_2 + 2nH_2O = (CH_2O)_n + nO_2 + nH_2O$ (24). The factors that control the rates of photosynthesis and thus, primary productivity are those that manage photosystems and their rates of reaction, as well as the rates of the dark reactions (24).

The Unites States Geological Survey lists the Mississippi River as the second largest river in the United States. The Environmental Protection Agency measures it to be 2,302 miles in length from its source, Lake Itasca in Clearwater Minnesota, to where it empties out into the Gulf of Mexico (43). As the world's 7^{th} largest river in its amount of discharge, an annual average of 10-35 x 10^3 m^3 s^{-1}, the Mississippi River is responsible for approximately 41 % of the drainage of the United States and is responsible for more that 70% of the freshwater input into the Gulf of Mexico (1, 42). Much of the drained land consists of farming areas and thus large amounts of nitrate and phosphorous, found in fertilizers, are deposited into the northern Gulf of Mexico. Turner and Rabalais (1991) estimated that 44% nitrogen and 28% phosphorous from the Mississippi River basin is deposited into the Gulf of Mexico by way of the Mississippi and Atchafalaya rivers. Additionally, it has been shown that the Mississippi River carries 111 ± 4.3 μg at NO^3-N 1^{-1} of nitrate and 7.4 ± 0.4 μg at P 1^{-1} pf phosphorous (42).

The increase in the amount of nitrate and phosphorous deposited into the Gulf has increased the amount of primary production occurring in the Mississippi River Plume (MRP) area (1). Studies have shown that primary production in the plume reaches as high as 8.17 g C m^{-2} d^{-1} (32). In a study to asses how the MRP affects the total surface-water production in the Gulf of Mexico, researchers calculated that the MRP was responsible for approximately 41% of all carbon fixation occurring in the upper 10 m of the oligotrophic water of the Gulf of Mexico (47).

1.2 The CBB Pathway and RuBisCO

There are several pathways utilized by other autotrophic organisms that allow them to fix CO_2, including the tricarboxylic acid pathway and the reductive acetyl-CoA pathway. However the main pathway by which photoautotrophs fix CO_2 is the Calvin-Benson-Bassham (CBB) pathway or reductive pentose phosphate pathway (40). The CBB pathway is the fundamental pathway for the movement of inorganic carbon in the biosphere and is conserved throughout evolution (40) (Fig. 1.1). This pathway consists of 13 enzymatic reactions and can be divided into two stages. In stage 1, three molecules of ribulose-1, 5-bisphosphate (RuBP) react with CO_2, catalyzed by ribulose-1, 5-bisphosphate carboxylase/oxygenase (RuBisCO), to give six molecules of 3-phosphoglycerate (PGA). One of the six molecules of PGA is used in the production of carbohydrates. In stage 2 the remaining five molecules of PGA are converted back into the starting substrate, RuBP through a series of reactions.

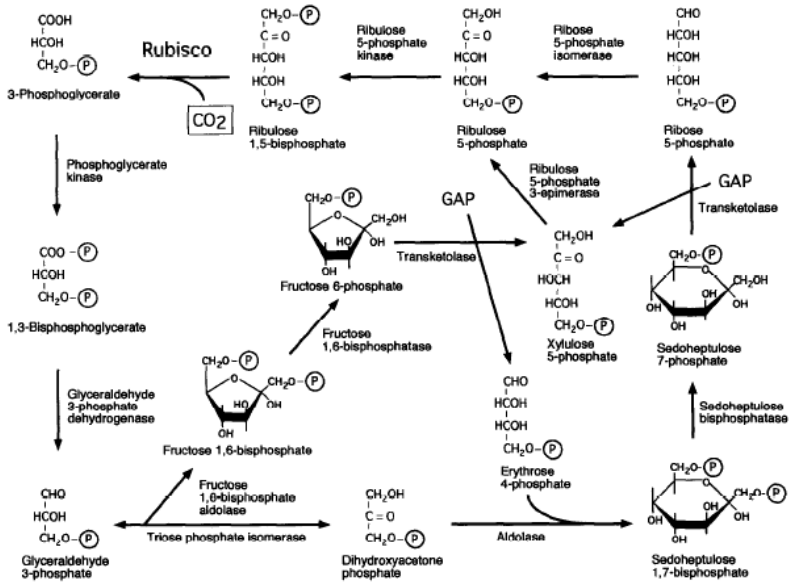


Figure 1.1. CBB Pathway. Figure taken from Atomi, 2002.

RuBisCO is the key enzyme in the CBB pathway (39, 40) because it is the first enzyme, of many, in the pathway and the only enzyme capable of fixing CO_2. The slow turnover rate of RuBisCO (1000-2000 mol CO_2 fixed/mol enzyme/min) and its poor catalytic activity (40), forces some plants to devote more than 50% of their protein content by weight to RuBisCO (15, 35) and thus it is considered to be the most abundant protein on Earth(11, 39, 40) . During photosynthesis, phytoplankton capture light energy from the sun. Carbon dioxide is taken up from the atmosphere and fixed via the CBB pathway. Carbohydrates are produced and oxygen is released into the atmosphere and surrounding waters. Carbon is incorporated into the cells of phytoplankton and cyanobacteria by either direct uptake of CO_2 or by the uptake of HCO_3 from surrounding waters (19).

RuBisCO has two functions: it catalyzes the oxygenolysis and carboxylation of RuBP (11, 40). CO_2 and O_2 actively compete with each other for the active site on

RuBisCO. In the alternate reaction, O_2 binds to the active site and is added to RuBP to yield phosphoglycolate which is metabolized in the glycolate pathway (35). This process, known as photorespiration, leads to a 50 % decrease in the overall efficiency of photosynthesis (35). The earliest forms of RuBisCO were not subject to this competition between CO_2 and O_2 because the atmospheric CO_2 concentration was 100 times higher and the O_2 concentration was less than 1 % of present day O_2 levels, therefore, if CO_2 entered the cell diffusively, the active site would be saturated with CO_2 (31). As the atmospheric concentrations of both O_2 and CO_2 changed over geological time, O_2 began to aggressively compete with CO_2 for the active site (31).

As a result of the competition between oxygen and carbon dioxide, many microorganisms have the ability to concentrate CO_2 at the carboxylation site (19). The development of carbon concentrating mechanisms (CCM) allow for these microorganisms to adapt to changing CO_2 levels in the atmospshere and ocean (19) as well as preventing oxygen from binding to the active site to form glycolic acid (40). Many species of cyanobacteria contain CCMs that enhance the efficiency of photosynthesis (3). Marcus et al. (1983) suggested that CCMs were induced by a product from the photorespiration pathway. When mentioning CCMs, one must take carboxysomes into consideration. Carboxysomes are cellular compartments that contain RuBisCO and are bound by a protein membrane (3, 22, 31). The polyhedral shape of these crystalline structures is similar to that of viral particles (3). It is believed that these structures are stores of RuBisCO that act as reservoirs to protect against photorespiration.

The functioning of CCMs in certain aquatic photoautotrophs is just now becoming understood (4, 14). The bulk of what is known about CCM functioning is based on a relatively limited number of organisms in culture, such as the chlorophyte *Chlamydomonas reinhardtii* (13, 37) and freshwater cyanobacteria *Synechococcus* sp. PCC7942 (51) and *Synechocystis* sp. Strain PCC6803 (23) among others. Such studies indicate that the components and genetics of CCMs are quite diverse, and some phytoplankton (i.e. the chrysophytes) appear to lack a CCM altogether, obtaining inorganic carbon (Ci) merely by diffusion.

One component of nearly all CCMs is the enzyme carbonic anhydrase. Carbonic anhydrases (CA) are a broad class of enzymes that catalyze the reversible conversion of bicarbonate (HCO_3^-) to CO_2. These may play a role in Ci uptake to the cell, as in the low-CO_2-inducible periplasmic CA of *C. reinhardtii* (37), or in facilitating the enrichment of CO_2 concentrations at the site of RuBisCO as for the carboxysomal CA of cyanobacteria (4). CAs are divided into 5 types (α, β, γ, δ, and ε) with no homology between them as they are thought to be the result of convergent evolution.

One subgroup among the putative CAs with greater conservation is the CsoS3 carboxysomal shell protein of marine α-cyanobacteria, a homologue of which has only recently been shown to have CA activity (36). No other putative carboxysomal CA gene is found in these organisms, as found for the β-cyanobacteria. Due to the critical role of CCMs in enabling phytoplankton to compete in highly productive coastal or freshwaters where Ci may be reduced, investigation of these mechanisms is warranted. Because of its relative conservation in sequence, the cyanobacterial CsoS3 genes represent a good target

for preliminary studies on detecting and quantifying the expression of CA genes in the environment and the relationship to changing CO_2 concentrations.

1.3 Forms and Structure of RuBisCO

RubisCO is found in several forms and is distinguished by subunit assembly and its biochemical properties (39, 48). There are four forms of Rubisco which include: Form I, Form II, Form III and Form IV. This review will focus on the major forms found in RubisCO containing organisms; Forms I and II (39). Form I consists of eight large (L_8) and eight small (S_8) subunits with a molecular weight around 550,000 Da (39, 45) and is found primarily in photosynthetic organisms and aerobic chemolithoautotrophs (12, 39, 45) (Fig. 1.2). This hexadecameric structure is the basic and most common of all of the forms and is conserved in many species of bacteria as well as higher plants (40). Form II RubisCO has only a large subunit and is found mainly in nonsulfur purple photosynthetic bacteria (12, 45) and marine dinoflagellates (26). These organisms usually fix CO_2 anaerobically (12).

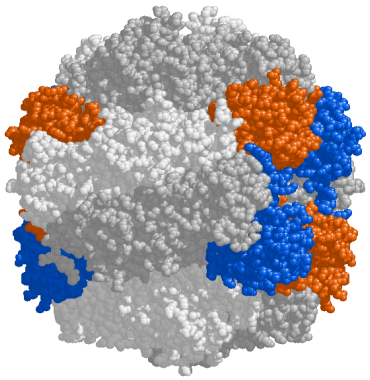


Figure 1.2. Quaternary structure of RubisCO. Image taken from Wikipedia.com (49)**.**

Form I is divided into subgroups denoted as “green” and “red” (12, 39). These subgroups are then further divided into subclasses: “green” is divided into IA and IB while “red” is divided into IC and ID (12, 39). RubisCO types from the green subgroup are found in green plants, green algae and cyanobacteria, while those from the red subgroup are found in red algae and purple bacteria (39). The picocyanobacteria *Prochlorococcus* and some *Synechococcus* species are the organisms in which most Form IA RubisCO is found (12, 48). Form IB is dominated by all of the green algae and other cyanobacteria (12, 48). Form ID RubisCO is found primarily in chromophytic, or nongreen, algae while Form IC has been found in alpha- and beta-proteobacteria (12, 48).

1.4 RuBisCO in the Marine Environment

The genes that encode for the large and small subunits of RubisCO are denoted as *rbcL*/*cbbL* and *rbcS*/*cbbS*, respectively (2, 40). It is on the large subunit that the active site for the carboxylation or oxygenation of RuBP is located (25, 40). If an organism expresses *rbcL*, then it is assumed that it produces RuBisCO and therefore fixes CO_2 by way of the CBB pathway. The phylogenetic patterns demonstrated by *rbcL* allow for researchers to identify the organism from which it came without the need for culturing the organism (44).

With the use of PCR, one of the first study of *rbcL* occurrence in a natural phytoplankton community was conducted (28). In this study, oligonucleotide primers were designed from sequences of a *Synechococcus sp*. to amplify conserved regions of the *rbcL* gene (28). These primers were then used in a PCR to DNA of phytoplankton

samples obtained from a Florida reservoir and algal isolates. Furthermore, extracellular DNA was also amplified thereby indicating that phytoplankton contribute to the fraction of dissolved DNA in the water column (28). Overall, this study was among the first to display the capabilities of PCR amplification to identify specific genes in natural populations.

rbcl transcription has been shown to occur on a diel cycle. Corredor et al. (2004) conducted a study on RuBisCO transcription and the photosynthetic capacity of phytoplankton. The researchers' objective was to determine whether the amount of mRNA transcription of a phytoplankton community, could be related to the biogeochemical cycles in which these communities play an integral role (9). Using the Geochemical Rate-RNA Integration Study, or GRIST, they observed that mRNA transcription and carbon fixation occurred on similar diel patterns and both reached their peak between 10 a.m. and 1 p.m. (9). In another study, researchers observed similar diel patterns in pure cultures of the cyanobacterium *Synechococcus* and the prymnesiophyte *Pavlova gyrans* (29). While both organisms were exposed to a 12 h light and 12 h dark cycle, both organisms exhibited slightly different transcription patterns. *rbcL* transcription levels in *Synechococcus* peaked around noon and then rapidly disappeared over the next 8 h. On the other hand, *P. gyrans's* transcription levels peaked at 4 p.m. and decreased to 66% of it's maximum level until 8 p.m. thus indicating that chromophytes are capable of fixing carbon later in the day than its cyanobacterium counterpart (29). Furthermore, Wyman et al. (52) examined the diel pattern in *rbcL* transcript abundance in a coccolithophorid. Results showed that the RuBisCO expression

peaked around sunrise and decreased by an order of magnitude later in the day. This report represents the first study of species-specific determinations of RuBisCO expression in a natural population as opposed to the aforementioned studies which deal with mixed or multiple populations.

1.5 Use of Array Technology in the Marine Environment

Macroarray and/or microarray technology has been used to analyze the function and occurrence of genes in the environment as well as the medical field. A microarray/macroarray works by exploiting the ability of a given nucleic acid molecule to bind specifically to, or hybridize to, a DNA template with which it shares homology. Arrays can be used to determine the expression of many genes in an environmental sample in one experiment. Generally, the process entails DNA in the form of oligonucleotides or amplicons from polymerase chain reaction (PCR) is spotted onto a glass slide or a nylon membrane. mRNA is then labeled or tagged with a fluorescent dye, either Cy3 or Cy5, and hybridized to the slide or membrane. Figure 1.3 shows a diagram of this process. The mRNA will bind to the strand of DNA most homologous in sequence to itself. The prefixes macro- and micro- refer to the size of the spot laid down on either the glass slide or nylon membrane (Fig. 1.4). Microarray spots must be viewed with high resolution cameras interfaced to special computer software and can hold several hundred of genes. Macroarrays are also analyzed with computer software but the spots can be seen with the naked eye and hold much fewer spots. Since many

microbial communities are unculturable, array technology can be an ideal tool in the identification of these communities.

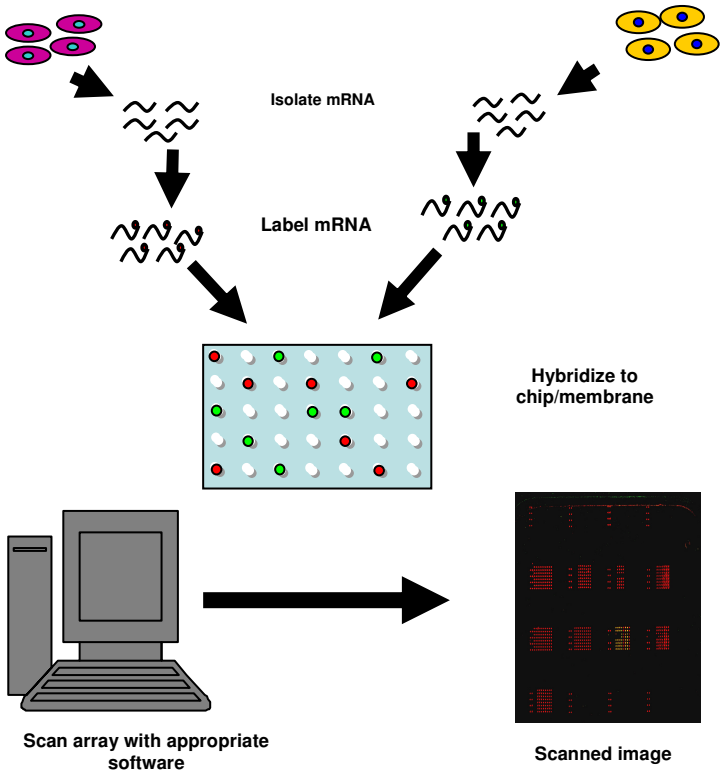
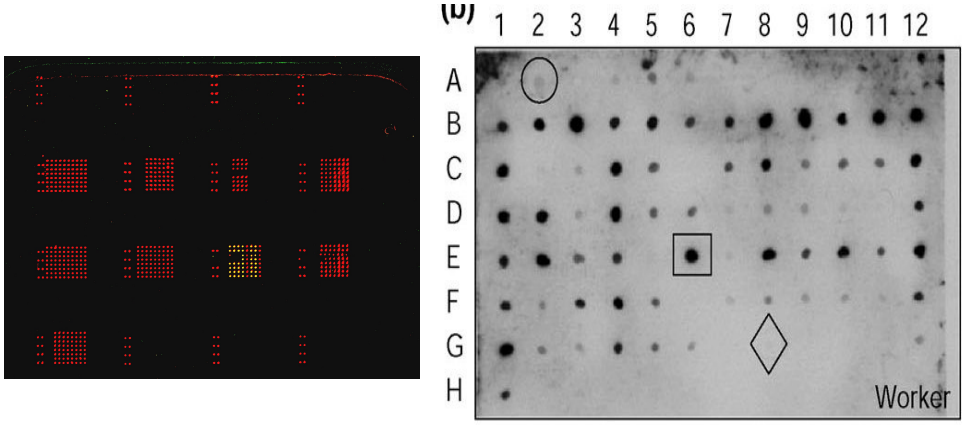


Figure 1.3. Microarry technology process. DNA is extracted from cells and transcribed into mRNA. Cells are labeled with a fluorescent dye and hybridized to a glass slide or memebrane. Array is then scanned to obtain an image and analyzed.

There are three general types of arrays. The first, and most common, type of array has different genes from a single organism that are spotted onto the glass slide or membrane in order to represent the genome of that organism (44). Arrays of this type are used to determine an organism's response to environmental changes. Another type of array utilizes the same gene obtained from different organisms. With this method of arraying, one is able to identify the organism from which each gene originated and is often used in diversity studies (44). The third type of array incorporates the use of

different functional genes from different organisms to quantify the amount of functional gene expression (44).



a) b)

Figure 1.4. Comparison of (a) microarray and (b) macroarray. Microarrays can hold more spots than the macroarray. Macroarray images courtesy Scharf et al. 2003.

Arrays have been used to determine the diversity of target genes in different environments. Jenkins et al. (2004) used macroarrys to fingerprint the diazotroph communities in the Chesapeake Bay. The results showed that Chesapeake Bay is home to a phylogenetically diverse diazotroph community (17). Macroarrays have been used to assess nitrogenase diversity in picoplankton. Steward et al. developed a macroarray to discern the capabilities of such technology to evaluate gene expression in the environment (38). They showed that macroarray results can be easily reproduced and are semiquantitative in assessing gene expression in a mixed sample (38).

Microarrays have also been used in microbial ecology studies in the identification and monitoring of bacterial communities found in wastewater and sludge. Loy et al. reports the use of a 16S rRNA oligonucleotide microarray, known as RHC-PhyloChip (probeBase), in the detection of bacteria of the order *Rhodocyclales* (21). Many of the

bacteria of this order are capable of degrading anthropogenic compounds. The results from this study showed that with the use of this specialized microarray allowed for the detection of *Rhodocyclales* population that represented less than 1% of the microbial community in the sample (21).

There are many obstacles that one must consider when constructing an array. When dealing with the same gene from different organisms, there is high probability that the sequences will be very similar to each other. This problem is not encountered in genomic arrays which target divergent genes from one organism. Another obstacle researchers may face is that when sampling from heterogeneous populations, probes may have very different melting temperatures (T_m) due to different sequence lengths (41). To correct for this, Taroncher-Oldenburg et al. (41) constructed oligonucleotides of equal length, minimal secondary structure and similar G-C content in the detection and quantification of functional genes found within the nitrogen cycle in the Chesapeake Bay system. The establishment of these parameters allowed for more uniform conditions during the hybridization process.

1.6 Real Time PCR in the Marine Environment

Real time PCR provides a means to observe the PCR while it is occurring. Unlike traditional PCR, real time PCR quantifies the amount of template produced in each cycle instead of at the end of the reaction. Real time PCR works by detecting the increase of fluorescence of DNA or RNA that has been bound to a fluorescent DNA stain, or to a fluorogenic probe specific for the target PCR product (48). SYBR green is a fluorescent

stain commonly used in real time PCR. The main disadvantage of SYBR green is that it has the tendency to produce false positives. TaqMan probes are fluorogenically labeled and utilize the 5` exonuclease activity of *Taq* DNA polymerase to provide a means of examining the amplification of specific PCR products.

Holland et al (1991) were the first to describe this process. They developed a method that utilized the 5` → 3` exonuclease activity of *Thermus aqaticus* (*Taq*) DNA polymerase in a PCR that allowed for the quantification of the target during each cycle of the amplification process (16). *Thermus aquaticus* is a bacterium found in hot pools that has revolutionized the way PCR is conducted today. *Taq* DNA polymerase replaced the Klenow fragment of *E. coli* DNA polymerase I PCR because its thermostability eliminates the need to add new DNA polymerase during every cycle of amplification.

In a study a study to examine *rbcL* expression in pelagophytes and diatoms, Wawrik et al. (2002), developed a method of detecting transcript abundance using real time PCR. A *Taq* Man probe specific for pelagophytes and diatoms was used in real time PCR and allowed for quantification of these groups in particular. Real time PCR data was compared to that obtained from ^{35}S- labeled oligonucleotide hybridization in order to determine the efficiency of the real time PCR method in microbial gene expression studies from environmental samples (48). The results from the study showed that the mRNA levels detected by both techniques were similar although hybridization levels were slightly higher (48). Higher hybridization mRNA expression levels were attributed to the predicted bias that hybridization experiments have towards degraded or partially degraded sequences (48).

The objective of this thesis was to determine what clades or phylogenetic groups are abundant in the primary production occurring in the Mississippi River Plume as determined by *rbcL* hybridization and real time PCR. I hypothesized that microarrays would play an important role in detecting the gene expression of *rbcL* in the MRP. The goals of my research were to first characterize the phytoplankton community in the MRP using microarray technology. The microarray data was then compared to RT-PCR and dot-blot hybridization data. Microarray data gathered for surface waters of the MRP showed high *Synechococcus*, pelagophyte and prymnesiophyte signals, although *Synechococcus* signals were at times significantly higher than those of the pelagopyte and prymnesiophyte probes. Dot-blot hybridization showed that the chromophytes dominated the surface by as much as 10 times the concentration of From IA groups. RT-PCR data from surface water showed that the Form ID organisms were 100 times more abundant than *Synechococcus*, suggesting that the strong signal of the *Synechococcus* probe on the array was a result of selective PCR amplification.

Chapter 2

Macroarray Detection of RubisCO Expression in Marine Environment

Chapter Summary

The Mississippi River is the 7th largest river in the world. It deposits a large amount of nutrients into the Gulf of Mexico (GOM) elevating nutrient levels in the Gulf and stimulating the growth of various phytoplankton species. The elevated nutrient levels in the Gulf causes an increase in the growth of various phytoplankton species. A DNA macroarray was constructed to determine its ability to quantify gene expression in the Mississippi River Plume (MRP). Ribulose- 1,5- bisphosphate carboxylase/oxygenase (RuBisCO) is the key enzyme in the Calvin-Benson-Bassham (CBB) pathway and the expression of the gene which encodes for this enzyme, rbcL, is indicative of carbon fization via the CBB pathway. PCR amplicons made from *rbcL* plasmid sequences obtained from the Mississippi River Plume were spotted onto a nylon membrane. The *rbcl* gene was excised from the plasmid, transcribed into RNA and then hybridized to the arrays. The macroarrays were not able to quantify gene expression in the MRP due to the frequency of cross hybridization of the probe to spots on the array from different phylogenetic groups.

2.1 Introduction

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) is the key enzyme in the Calvin-Benson-Bassham (CBB) pathway of photosynthetic carbon fixation in phytoplankton (40). The CBB pathway is the fundamental pathway for the entry of inorganic carbon into the biosphere (40). With regards to structure, RuBisCO is found in two major forms: Form I and Form II. Form I consists of eight large (L_8) and eight small (S_8) subunits. Form II consists of two large subunits (L_2). The genes that encode for the large and small subunits are denoted as *rbcL* and *rbcS*, respectively. The active site for carboxylation is found on the large subunit, and as a result, the expression of *rbcL* in an organism is indicative of carbon fixation via the CBB pathway.

Array technology provides a means of simultaneous identification and/or expression analysis of thousands of genes. Macroarrays and microarrays can be used in the detection of microbial strains and the presence of functional genes in an environmental sample. What is also unique about arrays is that the hybridization probe-target relationship is inverted. Traditionally, known DNA sequences are labeled with a fluorescent dye (Cy3 or Cy5), and hybridized to unknown samples on the slide. When using microarray technology, known DNA is spotted onto the slide and unknown samples are labeled and used as probes. DNA microarrays were originally developed for the study of nucleic acid sequences but now are commonly used in gene expression studies (20). Microarrays are ideal for use in the environment as most microorganisms are unculturable, they allow the identification and or functionality of genes found in the environment. Roszak and Colwell (1987) shed light on the viable but nonculturable

(VBNC) state in which most marine and estuarine bacteria exist. This term is given to microorganisms that cannot be maintained in a pure culture in the laboratory setting but do have detectable metabolic function (33). Array technology can be used to detect metabolic processes, such as the cycling of nutrients, occurring within these microbial communities by examining patterns of gene expression.

The Mississippi River is responsible for approximately 41% of the drainage of the United States and for more that 70 % of the freshwater input into the Gulf of Mexico (1, 42). Most of the land through which the Mississippi River passes is used primarily for agriculture. Phosphorous and nitrogen used in fertilizers are deposited into the river due to drainage from runoff and groundwater. These nutrients accumulate and are thusly deposited into the Gulf of Mexico. The Mississippi River Plume (MRP) is the dominant source of nutrients to the Gulf. This annual flux of nutrients causes massive phytoplankton blooms at mouth of the river, which extends out into the Gulf of Mexico.

The purpose of these experiments was to examine functional gene expression of ribulose-1,5-bisphosphate carboxylase/oxygenase in the MRP. The *rbcL* clades that will be used in this study were Form I type, specifically, clades IA, IB and ID. In this study, a macroarray designed to detect *rbcL* transcription was used to characterize the phytoplanktonic community in the oligotrophic Gulf of Mexico. Due to cross hybridization of the target sample to the probes on the arrays, macroarray results were inconclusive.

2.2 Methods

2.2.1 *Plasmid extraction*

Twenty-two clones known to already contain *rbcL* were selected from the work of Wawrik & Paul (47). These clones were obtained by amplification of the *rbcL* gene from natural phytoplankton communities of the Gulf of Mexico during the F.G. Walton Smith research cruise in the summer of 2001 (47). Figure 2.1 shows the cruise stations from which all samples were taken in the summer of 2001. Table 2.1 lists the clones that were used in this study as well as their corresponding clade. Clone names denote the cruise date, station, depth (A-H) from which it was obtained and primers used in amplification (Y for Form IA/IB; H for Form ID; and SY to denote when the cyanobacterial reverse primer was used). For example, WS01ST6CH17 is deciphered as Walton Smith, 2001, Station 6, depth C, primer H, clone 17 (47). Clone P994FY27 was taken from Pelican, 1999, Station 4, depth F, primer Y, clone 27 (46). All clones were kept at –80°C. Clones were streaked onto LB and Kanamycin plates and placed in the incubator at 37°C. Overnight cultures were prepared by taking one colony of each clone from LB and Kanamycin plates was put into 5 mL of LB broth and 5μL of Kanamycin (50 μg/μL). The cultures were placed in the shaking incubator at 37°C and 200 rpm overnight. Plasmid was extracted from each of the clones using the QIAprep Spin Miniprep Kit (Qiagen) and minipreps of each were sent off for sequencing at the University of Florida Core Sequencing Facility to ensure they were the correct clones. Upon clone sequence verification, clones were prepared for PCR.

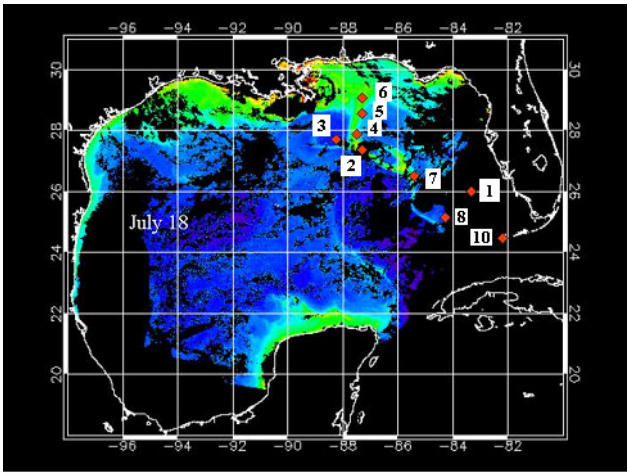


Figure 2.1. SeaWiFs chlorophyll satellite imagery of the Mississippi River Plume during time of sampling in 2001 overlaid with position of sampling stations.

2.2.2 PCR amplification and purification

Polymerase chain reaction was performed on each clone to amplify the *rbcL* target sequence. One µL of clone plasmid DNA was added to the PCR mixture that contained the primer set corresponding to its phylogenetic group. Form ID clones used 0.5 µL of 100 µM Form ID fwd (5'- GATGATGARAAYATTAACTC -3') and 0.5 µL of 100 µM Form ID rev (5'- ATTTGDCCACAGTGDATACCA -3') as the primer set (20 µM). Form IA and Form IB used 0.5 µL of 100 µM Syn fwd (5'- CTGAGCGGYAAGAACTAYGG -3') and 0.5 µL of 100 µM IA/B rev (5'- GGCATRTGCCANACGTGRAT –3') as the primer set (20 µM). In addition to the primer sets, the PCR mixture contained 1 µL of 10 mM deoxynucleoside triphosphates (800 µM), 5 µL thermophilic DNA polymerase 10X reaction buffer, 0.25 µL Taq DNA polymerase and 41.75 µL water. Thermocycler conditions consisted of an initial denaturation at 95°C for 2 min, followed by 40 cycles of 95°C for 1 min, 53°C for 1 min, and extension at 72°C for 1.5 min. There was an additional extension of 72°C for 5 min.

Amplicons were purified using the Zymo Kit (Zymo Research) and quantified using the Hoescht 33258 method (30).

Table 2.1. List of clone names, clade and common names.

Clone Name	Clade	Common Name
WS01ST6CH17	ID	Diatom
WS01ST4CH12	ID	Diatom
WS01ST4CH36	ID	Diatom
WS01ST6CH1	ID	Diatom
WS01ST7CH3	ID	Bolidomonas
P994CH1	ID	Pelagophyte
WS01ST7SY24	ID	Synechococcus
WS01ST8CH5	ID	Eustigmatophyte
WS01ST6CH33	ID	Xanthophyceae
WS01ST4CH16	ID	Dictophyceae
WS01ST8CH15	ID	Unknown, deeply rooted chromophyte
WS01ST1CH1	ID	Prymnesiophyte
WS01ST5CH10	ID	Prymnesiophyte
WS01ST2SY27	IA	Synechococcus
WS01ST2SY14	IA	Synechococcus
WS01ST2SY19	IA	Prochlorococcus
WS01ST3SY5	IA	Prochlorococcus
WS01ST8SY15	IB	Trichodesmium
WS01ST3SY4	IB	Chlorophyte
WS01ST6SY14	IB	Chlorophyte
P994FY27	IB	Prasinophytes
WS01ST8SY3	IB	Prasinophytes
WS01ST6SY8	IB	Prasinophytes

2.2.3 *Design of preliminary macroarrays*

After quantification of PCR amplicons, they were denatured with 10 mM RNAse-free EDTA and 0.4 M NaOH and then heated to 100°C for 10 minutes. An equal volume of 2 M ammonium acetate was added to neutralize the DNA mixture. Ammonium acetate acts as a buffer that neutralizes or lowers the pH of NaOH. Amplicons were dotted in duplicate onto a charged nylon membrane using a BioRad dot blotter. A diatom (WS01ST6CH1), prymnesiophyte (WS01ST5CH10), chlorophyte (WS01ST8SY13), chrysophyte (WS01ST1CH4), *Synechococcus* (WS01ST2SY27) and a prochlorophyte (WS01ST2SY19) were chosen to represent their clades. The amplicons were dotted at

different concentrations ranging from 100, 50, 10 and 1 ng per dot to determine which concentration is best for hybridization detection (see Figure 2.2). A total of six membranes were made.

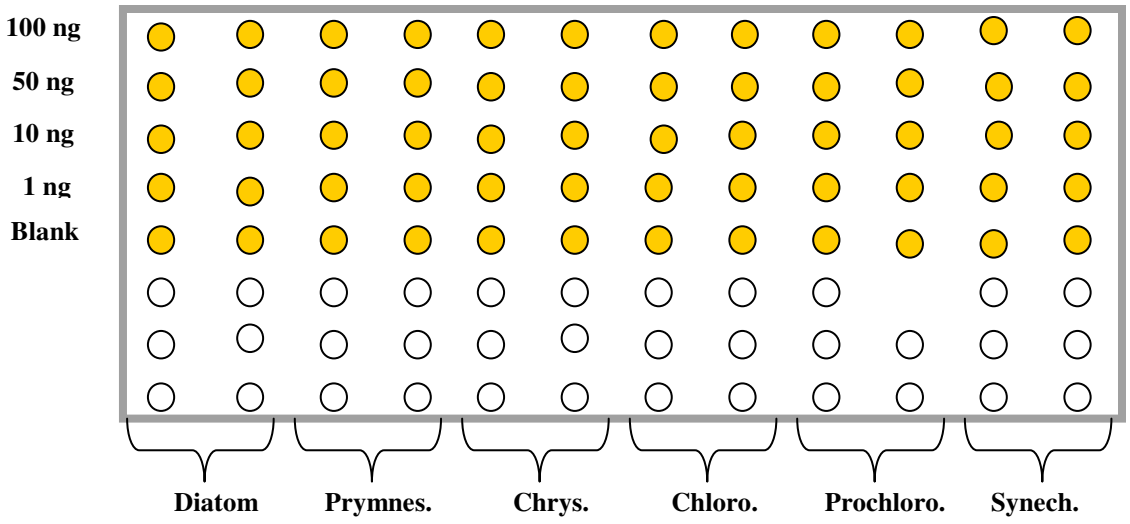


Figure 2.2. Preliminary blot schematic.
Clones were spotted onto the charged nylon membrane at the concentrations indicated.

2.2.4 *Transcript production*

Restriction digests were performed to linearize plasmid DNA in 40 µL reactions prior to in vitro transcription. Digestions were designed such that they would be at the 3' end of the sense orientation of the *rbcL* gene. The digest was purified by the Zymo DNA Purification Kit or the Promega SV Gel and PCR Purification Kit (Promega). The linearized plasmid DNA was then transcribed using the Riboprobe Combination System (Promega) for 2 h at 37°C using either the Sp6 or T7 RNA polymerase promoter to yield sense transcripts.

2.2.5 *Labeling of transcripts with biotin*

An annealing mixture was prepared containing 2.5 µg of transcript, 3 µL specific primer and RNAse-free water to a total volume of 10 µL. The annealing mixture was heated to 70°C for 3 min and then cooled to 42°C for 2 min. An RT Cocktail consisting of 5X GEAlabeling Buffer for Chemiluminescent Detection (BN) (SuperArray), Biotin, RNase-free water, RNase inhibitor and reverse transcriptase were combined and heated to 42°C for 1 minute before being added to the annealing mixture. The final mixture was incubated at 42°C for 90 min and then denatured at 94°C for 5 minutes. Before the labeled transcript was denatured, 1 µL of probe was removed and added to 19 µL of 1X TAE to give a 20-fold dilution. A 4-fold serial dilution was performed by taking 3 µL from the 20-fold dilution and adding it to 9 µL of 1X TAE. The remaining serial dilutions were 80-, 320-, 1280- and 5120-fold. This process was performed to assess the labeling efficiency of the probe.

2.2.6 *Hybridization and detection*

GEAhyb solution (SuperArray) was warmed to 60°C in a water bath. Sheared salmon sperm DNA (SuperArray) was heated to 100°C for 5 minutes and then immediately placed on ice. Arrays were prehybridized at 60°C for 2 h with 4 mL GEAhyb solution (Super Array) and 40 µL sheared salmon sperm DNA (Super Array). Another 4 mL of GEAhyb solution and 40 µL sheared salmon sperm DNA were incubated at 60°C in a hybridization oven. After prehybridization, prehyb solution was poured off. Four milliliters of hybridization solution and the entire volume of labeled transcript were

added to the hybridization tube. Each membrane was probed with a different transcript to ensure that the transcripts only hybridized to a member of their corresponding clade. Membranes were placed in the hybridization oven at 60°C with continuous rotation for 16 h. After hybridization, membranes were washed twice at 55°C for 15 min with a wash solution containing 2X SSC and 1% SDS with continuous rotation. Membranes were also washed twice at 55°C with a wash solution containing 0.1X SSC and 0.5% SDS at 55°C for 15 min with continuous rotation. These washing steps are a means of increasing the stringency of the array. The last wash was removed from the hybridization tubes and 5 mL of GEA Blocking Solution Q (SuperArray) were added and tubes were incubated for 40 min with continuous agitation at room temperature. After discarding GEAblocking Solution Q, 4 mL of Binding Buffer were added to the hybridization tubes. Tubes were incubated for 10 minutes with continuous agitation. Membranes were then washed four times with 8 mL 1X Buffer F (SuperArray) for 5 minutes each with gentle agitation. Next the membranes were rinsed twice with 6 mL Buffer G (SuperArray) for 5 minutes each. Arrays were then treated with CDP-Star for 2 h in the dark and then exposed to X-ray film for 20 min.

2.2.7 *Design of control macroarrys*

The control arrays were prepared in the same manner as the preliminary arrays, however, the full array was spotted with all 22 clones. Amplicons were dotted in duplicate onto a charged nylon membrane using a BioRad dot blotter. Each of the 22 clones were dotted in duplicate at 50 ng and 10 ng concentrations (Fig. 2.3).

2.2.8 *Transcription, hybridization and detection*

Four control arrays were tested with 4 clones that served as representations for their clades: a prasinophyte (8SY13), a prymnesiophyte (5CH10), a *Synechococcus* (2SY4) and a diatom (4CH12). Transcripts were made in the same manner as for the preliminary arrays. Control arrays were hybridized and detected in the same manner as the preliminary arrays.

50 ng	H_2O	6SY14	4CH16	8SY3	7CH3	7SY24	H_2O
10 ng	H_2O	6SY14	4CH16	8SY3	7CH3	7SY24	H_2O
50 ng	5CH10	8SY15	8CH5	4CH12	2SY19	4CH36	
10 ng	5CH10	8SY15	8CH5	4CH12	2SY19	4CH36	
50 ng	8SY13	8CH15	2SY4	6CH33	P994FY27	6CH1	
10 ng	8SY13	8CH15	2SY4	6CH33	P994FY27	6CH1	
50 ng	H_2O	6SY3	6CH17	5SY4	1CH4	P994CH1	H_2O
10 ng	H_2O	6SY3	6CH17	5SY4	1CH4	P994CH1	H_2O

Figure 2.3. Control blot schematic. Water was dotted in each corner to serve as a negative control. Clones were spotted in duplicate on the nylon membrane at 50ng and 10ng concentrations to determine which concentration gave the optimum spot density.

2.3 Results

The first few hybridization experiments were mostly unsuccessful due to the occurrence of cross hybridization of the transcript to the probe. However one experiment indicated that the transcripts were hybridizing properly. The blot shown in Figure 2.4 was hybridized with an 8SY13 transcript (Form IB) and strongly hybridized to itself. It also hybridized weakly to both 2SY19 (Form IA) and 2SY27 (Form IA). This hybridization pattern is to be expected since Form IA and Form IB *rbcLs* are closely related.

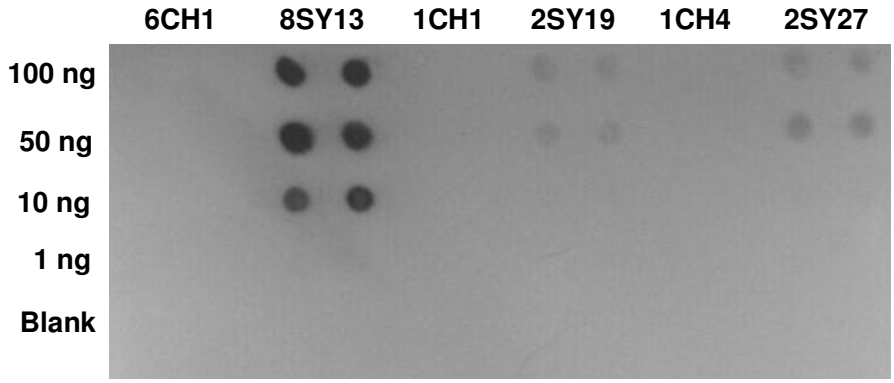


Figure 2.4. First successful hybridization experiment. The numbers to the left show the concentration of the dots laid down on the membrane. Notation across the top are the clones spotted onto the membrane: 6CH1- diatom; 8SY13- prasinophyte; 1CH1- chrysophyte; 2SY19- prochlorophyte,; 1CH4- prymnesiophyte; 2SY27- *Synechococcus*. This blot was hybridized with a transcript made from 8SY13.

Next the full array was spotted containing all 22 clones in the format described in Figure 2.3. Clones 1CH1 nad 2SY27 were replaced with 5CH10 and 2SY4, respectively. PCR amplicons were spotted in duplicate in concentrations of 50 ng and 10 ng. The first array was hybridized with 2.5 µg of an 8SY13 probe (Fig. 2.5 a). This experiment indicated cross hybridization occurring between unrelated clones. For example, on one array, a Form IB prasinophyte probe hybridized to a Form ID dictophyceae amplicon and

a Form IA *Synechococcus* amplicon. This had not been a problem in the preliminary arrays; as such the increase in cross hybridization led us to believe that the parameters for hybridization were not stringent enough. In an attempt to correct for this, the hybridization and washing temperature was increased from 60°C to 65°C. Increasing the hybridization and washing temperature did not prevent cross hybridization of the probes with dissimilar amplicons. Cross hybridization can be seen in Figures 2.5 b-c, which were hybridized at 65°C. Figures 2.5 b-c were hybridized with a diatom and a *Synechococcus sp*., respectively, and when compared to Figure 2.4, it is evident that cross hybridization has occurred. The diatom (Form ID) probe, hybridized to all *Synechococcus* dots and to all chlorophyte dots. Additionally, the *Synecococcus* probe hybridized with a pelagophyte, a dictophyceae, a prymnesiophyte all of which contain Form ID *rbcL*.

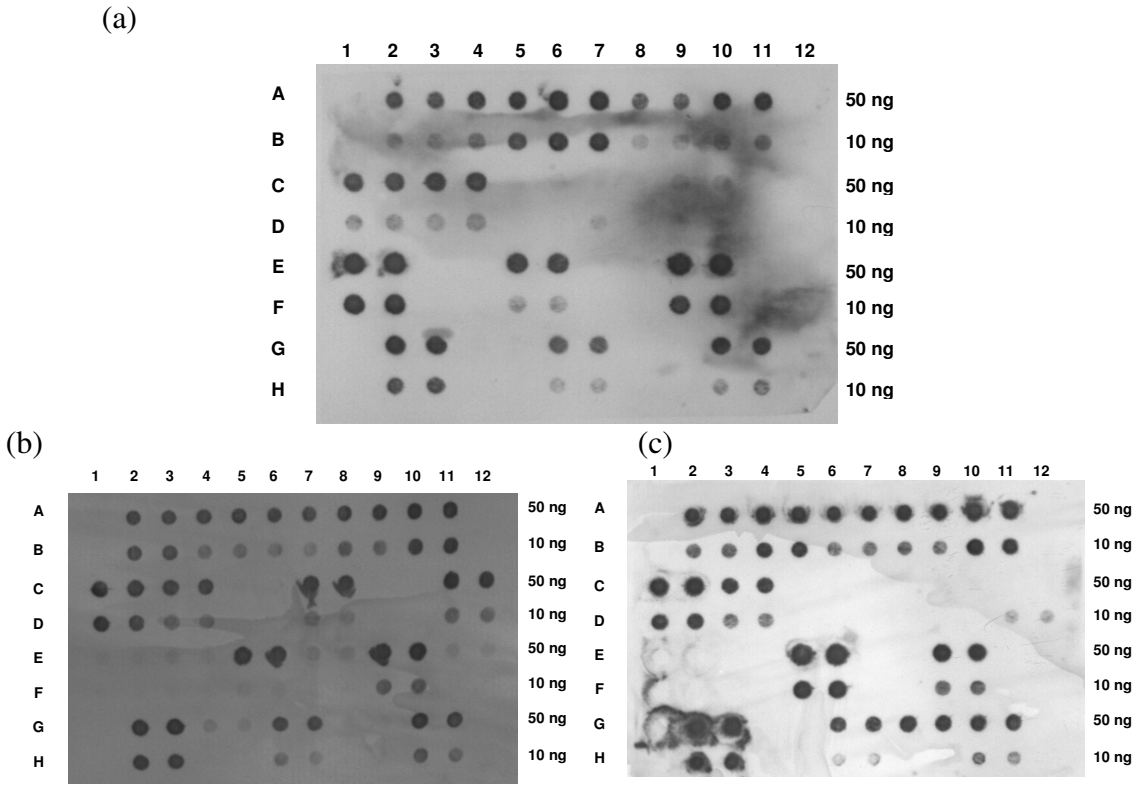


Figure 2.5 Fully dotted array experiments. (a) First hybridization experiment with the fully dotted array at 60°C. This array was probed with a labeled 8SY13 transcript (prasinophyte). (b) Array was hybridized with a 4CH12 (diatom) probe at 65°C. (c) Array was hybridized with a 2SY4 (*Synechococcus*) probe at 65°C.

2.4 Discussion

In this study a macroarray containing clones from the MRP was hybridized with transcripts from clones also obtained from the MRP. We were faced with difficulties from the onset of the experiment. Initially there was a problem with the labeling step using biotin. During the labeling efficiency step, there would be no spot on the film to indicate that the transcript was labeled with biotin. In an attempt to correct for this problem, many modifications were made to the protocol. We first assumed that the random primers provided by the labeling kit were not specific enough to our target and thus were not labeling the transcript. Specific primers were then used in place of the random primers. When changing the primers did not improve the labeling efficiency, we increased the amount of template by doubling the transcription reaction. We believe that the combination of the specific primer and the increase in the amount of template contributed to the success of the labeling step.

There was a considerable amount of cross hybridization occurring with the macroarrays. With the preliminary arrays there was also some cross hybridization but we believed it to errors made during the dotting of the macroarrays and thus decided to dot the full array. The non-specific binding of the target was still present; therefore the hybridization temperature was increased from 55°C, 60°C and 65°C, respectively, in separate reactions to increase the stringency of the array. Increasing the hybridization temperature did not reduce the occurrence of cross hybridization. One theory is that the initial concentration of the probe may have been too high and thus caused an overload of RNA to the array. During the labeling efficiency step, one must judge based on the

intensity of the spots as to what the optimal concentration for hybridization should be. A 1:10 dilution of probe 8SY13 was done prior to hybridization the array yet cross hybridization still occurred. In the next set of experiments, approximately 1.25 ng of labeled probe was used for hybridization (Fig 2.5 b-c). Even with half the original concentration of the probes, there was still significant cross hybridization with clones from unlike groups. Another possible cause of cross hybridization of the target could be a result of plasmid sequence overhang on the PCR products. The *rbcL* gene is excised from the plasmid by a restriction enzyme and transcribed into RNA. The PCR products that were spotted onto the array contained the plasmid sequence overhang. The pure transcripts also contain portions of the plasmid sequence which could result in the nonspecific binding of transcripts to spots from different phylogenetic groups.

Chapter 3

Microarray Detection of RuBisCO in the Mississippi River Plume

Chapter Summary

Here, a DNA microarray was used to detect *rbcL* expression occurring in the Mississippi River Plume (MRP). Total RNA was extracted from water samples collected from the MRP in 2005. cDNA was made from total RNA and amplified with *rbcL* primers in a polymerase chain reaction (PCR). PCR amplicons were labeled and hybridized to microarrays containing known PCR-amplified *rbcL* products obtained from various locations. All microarrays, constructed at Princeton University by the Ward lab, consisted of 70-mer oligonucleotides made from clones sequences previously obtained from the MRP. Information from the microarrays was compared to real-time PCR (RT-PCR) *rbcL* and dot-blot hybridization data obtained from the plume. Microarray data gathered for surface waters of the MRP showed high *Synechococcus*, pelagophyte and prymnesiophyte signals, although *Synechococcus* signals were at times significantly higher than those of the pelagopyte and prymnesiophyte probes. Subsurface data showed lesser amounts of *Synechococcus* and no *Prochlorococcus* expression. There were also significant signals observed in the chlorophyte probes. These results, provide a third layer of the detection of *rbcL* expression in the MRP.

3.1 Introduction

The Calvin-Benson-Bassham cycle is primary the pathway for the entry of inorganic carbon to the biosphere, and is the major pathway used by autotrophic organisms such as plants and phytoplankton. Ribulose-1,5- bisphosphate carboxylase oxygenase (RuBisCO) is the enzyme responsible for fixing CO_2 to ribulose bisphosphate. RuBisCO has four forms (I, II, III, and IV) that can be distinguished by their assemblage of subunits and their individual biochemical properties (39). Form I RuBisCO consists of eight large (L_8) and eight small (S_8) subunits. It can be further divided into two subgroups of either "green" or "red" (39). Each subgroup is then divided into subclasses: green is divided into IA and IB which are found in green plants, green algae and cyanobacteria; and red is divided into IC and ID which are found in red algae and purple bacteria (39).

Microarray technology offers a new approach to the study of microorganisms by providing a means to identify and examine the functionality of many genes simultaneously. Presently, majority of the studies using microarrays have been in the biomedical field, however, the use of microarrays in environmental studies has begun to increase. The lack of environmental studies is widely due to difficulties associated with cultivating microbial communities found in environmental samples. Nonspecific probe to target binging, temperature variability amongst probes and problems with DNA amplification are the main obstacles that must be addressed when working with microarrays (41).

The Mississippi River is responsible for over 41% of the drainage of the United States and thus deposits a significant amount of nutrients into the Gulf of Mexico (1, 42) forming a plume. These nutrient deposits fuel annual phytoplankton blooms. Past studies have shown the presence of diverse phytoplankton groups in the plume (46). In this study, *rbcL* obtained from the Mississippi River Plume (MRP) was extracted and hybridized to a microarray containing known *rbcL* probes. This data was then compared to real time PCR and dot-blot hybridization data obtained from the plume.

3.2 Methods

3.2.1 *RNA extraction*

Three-tenths to 20 L of water collected from the Mississippi River Plume in July 2005 were filtered with sterivex filters and 350 µL of RLT buffer were added to each filter. Three to nine-hundred mL of water were also filtered with 0.45 µm Durapore filters and which were placed in 2 mL tubes with 750 µL RLT buffer and silica or glass beads. Filters were placed in the -80°C freezer for storage. Total RNA was extracted from the filters following the RNeasy Mini Kit (Qiagen) protocol. Table 3.1 lists the cruise samples that were used. Figure 3.1 shows the cruise stations from which these samples were collected.

Table 3.1. List of cruise samples showing depth, amount filtered and type of filter used.

Station	Depth	Water Filtered (L)	Filter
8B	55m	15	Sterivex
7X	3m	11.2	Sterivex
TB	3m	0.3	Durapore
2G	100m	2	Sterivex
6A	3m	2.6	Sterivex
5A	3m	2.6	Sterivex
1B	60m	15	Sterivex

3.2.2 *Transcription*

cDNA was made using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) according to the manufacturer's protocol. Briefly, total RNA up to 5 µg, but no less than 1 pg, was combined with 50 ng/uL random hexamers and a 10 mM dNTP mixture. Samples were incubated at 65°C for 5 minutes and then placed on ice. cDNA Synthesis mix was prepared using 10X RT buffer, 25 mM $MgCl_2$, 0.1 M DTT, RNaseOut

(40 U/µL) and SuperScript III RT (200 U/µL) and added to the RNA/primer mixture. This new mixture was then incubated first at 25°C for 10 minutes and then at 50°C for 50 minutes followed by incubation at 85°C for 5 minutes to terminate the reaction. Samples were placed on ice following incubations and RNAse H reagent was added to each sample followed by incubation at 37°C for 20 minutes. cDNA was then quantified using the Quant-iT Pico Green dsDNA Assay Kit (Invitrogen). cDNA was amplified with PCR using all three *rbcL* primer sets (Form ID fwd and rev; Form IA/IB fwd and rev; Syn fwd, Form IA/IB rev) before labeling. cDNA ranging in concentration from 0.1 to 2 ng was added to a mixture containg the forward and reverse primer (100 µM each) and PCR Master Mix (Promega). Thermocycler conditions consisted of an initial denaturation at 95°C for 2 min, followed by 40 cycles of 95°C for 1 min, 52°C for 1 min, and extension at 72°C for 1.5 min. There was an additional extension of 72°C for 5 min. Amplicons were purified using the Zymo Kit and quantified with the Quant-iT Pico Green dsDNA Assay Kit (Invitrogen).

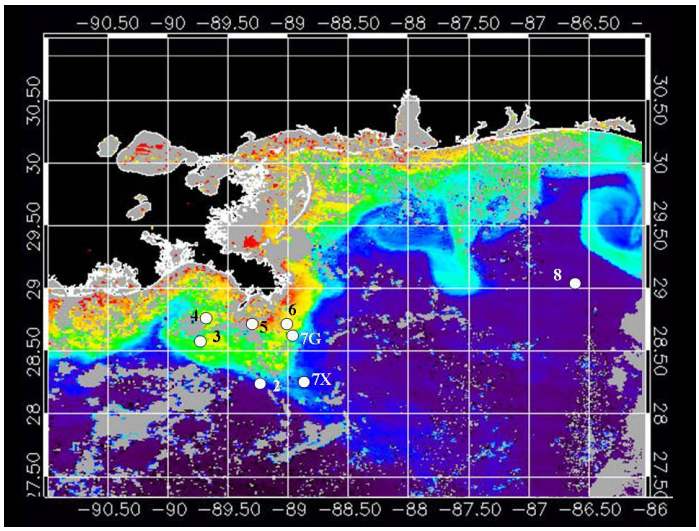


Figure 3.1. SeaWiFS ocean color satellite image of the Mississippi River Plume overlaid with cruise stations.

3.2.3 *Positive controls*

Transcripts of clones obtained from the Mississippi River Plume in 2001 (46, 47) were made first by digesting plamids containing the *rbcL* insert with the appropriate enzymes. Table 3.2 lists the clones that were used as positive controls. All digests were purified with the Zymo DNA Purification Kit (Zymo Research). The linearized plasmid DNA was then transcribed using the Riboprobe Combination System (Promega) for 2 h at 37°C using either the Sp6 or T7 RNA polymerase promoter to yield sense transcripts.

Table 3.2. List of the transcripts used as positive controls for the microarray performance tests.

Clone	Family	RuBisCO Type	Reference
P994CH1	Pelagophyte	ID	Wawrik et al., 2003
WS01ST1CH1	Prymnesiophyte	ID	Wawrik & Paul, 2004
WS01ST8CH12	Chrysophyte	ID	Wawrik & Paul, 2004
WS01ST3SY1	HL *Prochlorococcus*	IA	Wawrik & Paul, 2004
P994GY7	LL *Prochlorococcus*	IA	Wawrik et al., 2003
WS01ST6SY3	*Synechococcus*	IA	Wawrik & Paul, 2004
WS01ST4SY39	Prasinophyte	IB	Wawrik & Paul, 2004

3.2.4 *Labeling*

Amplified cDNA was combined in order to minimize within reaction variability and bias. Mixed cDNA was labeled in a random priming labeling reaction using the BioPrime Array CGH Genomic Labeling Module (Invitrogen) with slight modifications. In short, a reaction mixture containing 2.5X random primers (125 mM Tris-HCl, template cDNA, 12 mM $MgCl_2$, 25 mM 2-mercaptoethanol, 750 µg/mL oligodeoxyribonucleotide primers), 1.2 mM dNTP mix with aminoallyl-dUTP (10 mM dACG mix, 10 mM dTTP, 10 mM dUaa) and Klenow enzyme (40 U/L Klenow fragment in 100 mM KPO_4, 1 mM DTT and 50% glycerol). This mixture will be incubated at 37°C for 2 h. Labeled product was cleaned with the Qiaquick PCR Purification kit (Qiagen) and quantified with

the Quant-iT Pico Green dsDNA Assay Kit (Invitrogen) to ensure that the labeling procedure was successful. Following quantification, the labeled product was dried down to a pellet and frozen at -80°C.

3.2.5 *Coupling of dUTPaa-labeled target to Cy3 dye*

First, the Cy3 dye pellet was re-suspended in 40 µL of dimethyl sulfoxide (DMSO). The dUaa labeled pellet was thawed and re-suspended in 4.5 µL of 0.10 M Na_2CO_3 buffer and incubated at room temperature for 15 min. Next a 4.5 µL aliquot of dye solution was added to the re-suspended pellet and allowed to incubate in the dark for 1 hr. Once incubation is complete, 4.5 µL 4M hydroxylamine were added to the mixture and allowed to incubate in the dark for an additional 15 min. Hydroxylamine quenches the coupling reaction of the Cy3 dye with the dUaa label. The mixture was then purified with the Qiaquick PCR purification kit with minor modifications: 25 µL of ddH_2O were added to each sample before the addition of Buffer PB; 3 µL 3M NaOAc was added to ensure low pH of Buffer PB; samples were washed 5 times with Buffer PE; 30 µL Buffer EB was added and columns were allowed to sit for 2 min, and then spun down in a microcentrifuge. The elution step was repeated to yield 60 µL of target. Finally, labeled samples were quantified with the Pico Green kit for double stranded DNA and dried down into a pellet.

3.2.6 *Microarray construction and hybridization*

Two types of slides were used in this study (Fig. 3.2). The first type, referenced as BC008 (Fig 3.2a), consisted of a glass slide and a single gasket. Only one probe can be used with this array. The second slide, referenced as BC009 (Fig 3.2b), is unique in that

there are two gaskets. This allows for the hybridization of two probe sets in a single experiment in which a different hybridization solution, each with a different Cy3 target, is placed into each gasket and the array slide is placed onto the coverslip. The DNA laid down on the array were 90-mer oligonucleotides from various organisms in which the first 20 bases were complimentary to the Cy5 reference oligonucleotide used in the hybridization solution (Table A-1). All microarray slides were printed at the Microarray Facility at Princeton University Princeton, New Jersey. This is an ozone free facility due to the sensitivity of Cy5 to the gas.

The hybridization protocol for BC009 is different from BC008. The hybridization solution consisted of 100 µL 2X Hybridization Buffer (Agilient), (*n*) µL Cy3-target, (*n*) µL Cy5-reference oligonucleotides and distilled water to a total of 200 µL in a 1.5 mL tube. A total of 4 ng of transcript cDNA and 100 ng of cruise sample cDNA were used in the hybridization solutions. Tubes were mixed and heated to 95°C in a wet block with the lid locked for 5 min and then allowed to cool to room temperature for a minimum of 2 min. The hybridization solution was then applied to the gasket of the slides. Slides were incubated in a rotating oven overnight at 60°C.

When using the BC008 slides, 30 mL of prehybridization solution was made containing 7.5 mL of 20X SSC (5X in 30 mL), 3 mL of 10 % SDS(1 % in 30 mL) and 0.3g (1 % in 30 mL) of a blocking reagent in the form of chicken or bovine serum albumen (CSA or BSA). This solution was heated in 5 s intervals until clear and all CSA/BSA is melted. It was then filtered through a 0.45 µm filter. Array slides were incubated in this solution for a minimum of 40 min at 64°C in the hybridization oven or

warm water bath. After incubation, slides were dipped in e-pure water 5 times and then rinsed with isopropyl alcohol. Slides were then centrifuged to remove any residual alcohol.

Next, parafilm was wrapped around the slides to serve as a barrier between the coverslip and the array slide. Slides were then placed label side up in the hybridization chamber. The coverslip was secured with more parafilm and the slides were placed into the hybridization chamber. Hybridization solution containing 200 ng target DNA, 1 μL Cy5 reference oligonucleotide and prehybridization solution to a total of 80 μL was pipetted onto the array slide under the coverslip. The hybridization chamber was sealed with black clamps and wrapped in foil to block out any light. Chambers were then placed in the warm water bath or hybridization oven and incubated overnight.

3.2.7 *Washing and Scanning*

BC009 slides were removed from the rotating oven and placed in Wash #1(20X SSC, 10% SDS, water) and shaken for 10 min at 100 rpm. The slides were then removed from Wash #1 and placed in Wash #2 (20X SSC, water) for 10 min at 100 rpm. For the final wash, Wash #3 (20X SSC), slides were also shaken for 10 min at 100 rpm. BC008 slides use the same reagents, time and rpm's in the washing steps but the concentrations are slightly different. Wash #1 contains 1X SSC and 0.05 % SDS at 55C; Wash #2 contains 0.1X SSC and 0.05 % SDS; and Wash #3 contains 0.1X SSC. All arrays were scanned in the Agilent Scanner (Agilent Technologies) and analyzed with Gene Pix Software (Molecular Devices).

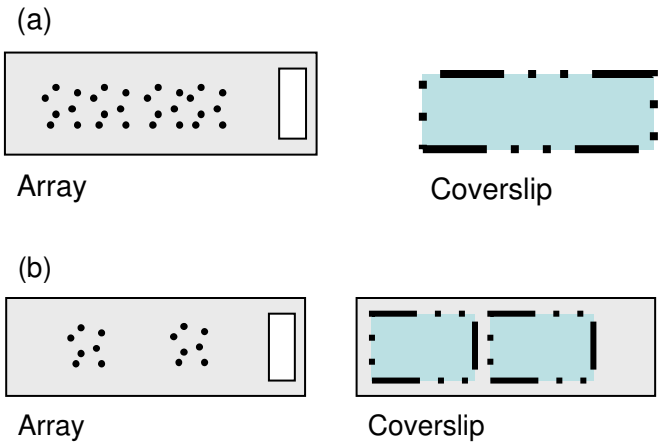


Figure 3.2. Types of array slides. The first type (a), referred to as BC008, consisted of a glass slide with one gasket in which only one probe set can be used. The second type (b), referred to as BC009, consisted of a glass slide and a coverslip with two gaskets. This allows for hybridization of a slide with two different probe sets.

3.3 Results

cDNA made from cruise samples was sent to the Ward lab in July 2005 and the remaining cDNA was sent in November 2006. Many of the probes on the array consisted of a group of *rbcL* sequences obtained from GenBank, from which a representative sequence based on similarities in that group, was selected to characterize all types of that particular group. The probes with accession numbers were also obtained from GenBank. Table A.1 lists the probes and the representative sequence that was chosen. All data were normalized by dividing all Cy3/Cy5 ratios for each probe, by the highest ratio in the data set.

3.3.1 *Microarray performance*

Microarray data was arranged into groups: controls, surface samples and subsurface samples. Transcript DNA from clones obtained from the MRP in 2001 (46, 47) were hybridized to microarrays. Figures 3.3-3.6 show the transcripts that were used to test the performance of the microarrays. Clone 3SY1 is a high-light *Prochlorococcus* (Fig. 3.3); clone 6SY3 is a marine type A *Synechococcus* (Fig 3.4); clones P994CH1 and P994GY7 are a pelagophyte and low-light *Prochlorococcus*, respectively (Fig. 3.5); and clones 8CH12 and 1CH1 are a chrysophyte and a prymnesiophyte, respectively (Fig. 3.6). All of the clones, with the exception of 8CH12 and P994GY7, hybridized to the probe whose sequence to which it was most homologous. Clone 8CH12 did not hybridize because there was no similar target on the array.

Since P994GY7 did not hybridize specifically to one probe on the array, a pairwised sequence alignment was performed to determine how similar the sequences

were to the respective probes to which each hybridized. The sequence alignment of P994GY7 showed that it is at least 68% similar to AF381708, an uncultured chlorophyte and to *Prochlorococcus marinus*. The probe with the largest signal was the low light *Prochlorococcus* consensus probe and thus was expected to show a stronger signal than the other probes.

3.3.2 *Tampa Bay*

Fort Desoto is located at the entrance of Tampa Bay in the southwest corner of Mullet Key. The water samples from Fort Desoto were taken at 3m depth (Table 3.1). There was a large amount of background signal in the Fort Desoto hybridization experiment (Fig. 3.7). This array was hybridized with *rbcL* product made with Form ID primers, thus only the last eight probes on the far left should have a signal. The signals of the Form IA and Form IB probes were caused by a large background not observed in other samples. Prymnesiophytes were dominant at Fort Desoto, followed by the pelagophytes, the silicoflagellates and the haptophytes. There were also significant amounts of *Phaeodactylum tricornutm*, *Karenia mikimotoi*, diatoms and phaeophytes. Figure 3.8 shows another hybridization experiment with cDNA from Fort Desoto amplified with Form ID primers. In this experiment there is a high signal from the prymnesiophytes. The remaining Form ID probes were below the Cy3/Cy5 ratio cut off. This array was hybridized with MRP Station 2G amplified with Form IA/IB primers, which is discussed later.

3.3.3 *Surface samples*

The surface MRP samples were obtained from Stations 5A and 6A. Water samples from Station 5A were taken at 3m and a total of 3.6 L of water was filtered with Sterivex filters (Table 3.1). Samples collected from Station 6A were taken at 3 m depth and 2.6 L were filtered with Sterviex filters (see Table 3.1). Figure 3.9 shows the hybridization data from Station 5A for the cDNA that was sent in July 2005 to the Ward lab in Princeton. Figure 3.10 shows the data from Station 5A for the arrays that were hybridized in November 2006. cDNA used in Experiments 5A-1 and 6A-1 were amplified with only Form ID primers. cDNA used in Experiments 5A-2 and 6A-2 were amplified with both Form ID and Form IA/IB primers. In Experiment 5A-1, pelagophytes are the dominant group followed by diatoms, prymnesiophytes, *P. tricornutm*, haptophytes and silicoflagellates. There are small signals in some of the Form IA and Form IB probes but because only Form ID primers were used, these signals are attributed to background. Although the Form ID probes of array 5A-2 appear to have weaker signals than those observed in Station 5A-1, they are proportionate to each other and normalized to the *Synechococcus* signal. *Synechococcus* gave the highest signal at Station 5A-2. In array 6A-1, pelagophytes and prymnesiophytes dominated amongst the Form ID probes (Fig. 3.11). There were also significant signals for the haptophytes and silicoflagellates. *P. tricornutum*, pelagophytes and prymnesiophytes dominated array 6A-2 (Fig. 3.12). There was also a strong signal in the *Synechococcus* probe. As shown in array Station 5A-2, the Form ID probes in Station 6A-2 were lower than those of Station 6A-1, however the

same proportions are observed. Small amount of chlorophytes were present at both Stations 5A-2 and 6A-2 (Figs. 3.12-3.14).

3.3.4 *Subsurface samples*

Subsurface samples were obtained from Stations 8B, 1B, 2G and 7X (Fig. 3.1). In all four stations subsurface stations there is a similar distribution of Form ID containing organisms. Station 8B is located at the subsurface chlorophyll maximum (SCM) and not in the plume. It is one of only two subsurface stations in which there is data available for Form ID and Form IA/IB (Fig. 3.13). Water samples collected from Station 8B were obtained at 55 m (Table 3.1). Pelagophytes dominated the Form ID probes and there was a significant signal from the prymnesiophytes. Amongst the Form IA/IB probes, *Synechococcus* was the dominant player at Station 8B but there was also a significant signal from the prasinophyte, Chlorella, and the uncultivated chlorophyte AF381699. Thus, this subsurface sample was the first to demonstrate appreciable hybridization to chlorophyte (green algal) phytoplankton.

The SCM of Station 1 (1B) was also sampled (Fig. 3.14). These samples were taken from a depth of 60 m of which 15 L of water were filtered. This array was hybridized with prasinophyte transcript (4SY39) and cDNA obtained from Station 1B. Pelagophytes yielded twice the signal of prymnesiophytes and there were significant amouns of haptophytes, silicoflagellates, diatoms and *K. mikimotoi*. Transcript 4SY39 did not hybridize to any of the chlorophyte/prasinophyte probes due to poor labeling efficiency.

Station 2G is located just outside of the plume at a depth of 100m. cDNA from this station was amplified with both Form ID and Form IA/IB primers, however on different arrays. Figure 3.15 shows the cDNA that was amplified with Form ID primers and as expected, only the Form ID probes have signals. Pelagophytes are once again the dominant player at Station 2G followed by the prymnesiophytes. There are also significant signals for the haptophytes, silicoflagellates and diatoms. Figure 3.8 shows cDNA from Station 2G amplified with Form IA/IB primers and aforementioned cDNA obtained from Fort Desoto amplified with Form ID primers. *Synechococcus* dominated the Form IA/IB probes. *Chlorella* and the prasinophytes were significant at Station 2G as well as a few of the other chlorophyte probes, as was observed for the other subsurface sample, 8B. Station 7X is located outside of the plume at a depth of 3m (see Fig. 3.16). Pelagophytes dominated among the Form ID probes followed by significant signals from some of the other Form ID probes. This array was also hybridized with a *Synechococcus* transcript and thus explains the strong signal with *Synechococcus* consensus probe.

3.4 Discussion

A microarray was designed to ascertain which clades or phylogenetic groups are most abundant and active in the primary production occurring in the MRP as determined by *rbcL* hybridization and real time PCR. Microarrays were designed by the Ward Lab Group at Princeton University, Princeton, NJ. The probes used on the array are made up of a group of sequences taken from GenBank that were then assembled into consensus groups. Hybridization experiments were separated into three groups: controls, surface and subsurface.

With the exception of the prasinophyte and the chrysophyte, all positive controls hybridized to their respective probes on the array. Initially, we were unaware of what was exactly on the array and a chrysophyte was chosen as a positive control due to its presence in the MRP. However there was no chrysophyte probe on the array and thus no hybridization signal was observed. The prasinophyte probe did not label due to poor labeling efficiency of the Cy3 dye the dUTPaa labeled target.

cDNA from Fort Desoto was amplified with only Form ID primers, yet there was a signal for every probe on the array including the Form IA and Form IB probes (Figure 3.7). The background levels in this experiment were particularly high. Since neither Form IA nor Form IB primers were used to amplify the cDNA, these signals were ignored and only the Form ID signals were analyzed. These results are in agreement with real time PCR data obtained from Tampa Bay. The real time PCR assay showed that haptophyte *rbcL* expression, which includes prymnesiophytes, was significant in Tampa Bay where they were found in concentrations as high as 39 pg L^{-1}. Furthermore,

heterkont *rbcL* expression, which includes diatoms, pelagophytes, pinguiphytes and some silicoflagellates was found in concentrations as high as 402 pg L^{-1}.

Surface stations had high *Synechococcus*, pelagophyte and prymnesiophyte signals. There were no significant signals observed in the chlorophte or *Prochlorococcus* probes. We were able to obtain phylogenetic profiles for both Form ID and Form IA/IB groups for Stations 5A and 6A. Dot blot hybridization showed that the chromophytes dominated the surface waters by as much as 10 times the concentration of Form IA groups (18). Furthermore, RT-PCR results from Stations 5A and 6A showed that the heterokonts were 100 times more abundant than *Synechococcus* (18). The strong signal of the *Synechococcus* probe could be due to selective PCR amplification and not due to an abundance of *Synechococcus* in the plume surface. However, virtually no signal was obtained by PCR or microarray analysis for any *Prochlorococcus* phylotypes in the surface waters of the MRP.

Stations 5A and 6A were located within the plume where it has been previously shown that diatoms were dominant among the microplankton in the plume (18). Microarray analysis revealed an abundance of pelagophytes in the plume. Since the real time PCR assay cannot distinguish between diatoms and pelagophytes given their close relationship to each other (46), it is probable that the abundance of diatoms previously observed is actually a mixture of diatoms and pelagophytes. When Station 5A-2 and 6A-2 results were compared to Stations 5A and 6A, it was evident that there was a difference in chromophyte signals on the arrays. However due to the high *Synechococcus* probe signal, the Form ID probe signals are normalized to the Synechococcus signal and appear

weaker than those observed in arrays 5A-2 and 6A-2. Nonetheless, a similar distribution of organisms is still observed across the two surface stations.

Real time PCR assays conducted by John et al. (Submitted) indicated that diatoms were more than likely responsible for most of the rbcL expression occurring in the MRP. This coincides with microarray analysis which showed that Form ID organisms, specifically the pelagophytes and prymnesiophytes, were in greater abundance in the surface and subsurface waters. Subsurface samples had lesser amounts of *Synechococcus* than the surface stations and no prochlorophytes. Pelagophytes were dominant along the SCM. Real time PCR and dot blot hybridization resulted in significant amounts of heterokonts in the subsurface. Station 8B was the first station at which signals were observed from the Form IB clade. Wawrik et al., (2003) also found that prasinophytes were high at the SCM and the presence of diatoms.

Microarrays detected no chlorophytes in the surface but they were found at the SCM. John et al. (Submitted) also found no chlorophytes in the surface. Station 2G was dominated by Form ID *rbcL* expression, but there were also significant amounts of *Synechococcus* and *Chlorella rbcL* expression. This data coincides with the work of Wawrik et al. (2003) in which a clone library was constructed containing a diverse group of phylotypes closely related to green algae. One of the predominant groups among these clones was closely related to *Chlorella* sp. (46). This finding agrees with the *Chlorella* probe signal on the microarray (Fig. 3.8). The other group discovered was related to *Bathycoccus prasinos* and was abundant at the SCM (46). A significant prasinophyte signal was observed in the microarrays at Station 8B (Figure 3.13) which is located at the

SCM. There was a significant pelagophyte signal at Station 7X. This station was located near the surface where our data shows that pelagophytes dominated other surface stations.

Previous studies have used flow cytometry and epifluorescence microscopy to study picoplankton community structure in the ocean (5, 6) and thus it is believed that in tropical and subtropical areas, *Prochlorococcus* dominates the picoplankton. Wawrik et al. (2003) found an abundance of *Prochlorococcus* in the middle water column of the MRP around 40 m depth. However, no *Prochlorococcus* probe signals were observed in the subsurface stations on the microarrays. Additionally there were no Prochlorococcus signals in the surface waters of the MRP which supports the work of Wawrik et al., (2003) in which *Synechococcus* dominated the surface waters. Usually, Synechococcus dominates where *Prochlorococcus* is found in lesser amounts and thus could explain why Prochlorococcus was not found in the subsurface. Abundance in *Synechococcus* has previously been shown to occur at the interface where plume waters and water form the Gulf of Mexico mix (47). Furthermore, *Prochlorococcus* is abundant outside of plume waters in the oligotrophic waters of the Gulf of Mexico. Of the four stations with high to significant signals for the *Synechococcus* probe, Station 2G was the only one located at the interface where plume waters and blue waters meet.

This array design does not provide a thorough representation of the phylogenetic diversity of phytoplankton in the MRP. Figure A-3 in the Appendix shows the phylogenetic tree of the clone library made by Wawrik et al. (46). Table A.1 lists the probes used on the microarray and their sequences. When the list of probes is compared

to the phylogenetic tree, it is evident that many of the sequences used in different consensus groups overlap one another on the tree (Figs. A-1 – A-3). The BroadChlr1 consensus group encompasses most of the BroadChlr2 consensus and probe AF381699. Furthermore, BroadChlr1 consensus group contains a few *Chlorella* species. All of the clones used in the haptophyte consensus were classified as prymnesiophytes by Wawrik et al. (2003) and only one prymnesiophyte sequence was used for the prymnesiophyte probe. Figures A-4 – A-6 show updated and more detailed trees containing the sequences used to make the probe consensus groups, the transcripts used as controls and their closet relatives.

This study provides a third level of specificity at which phylogenetic diversity has been sampled in the MRP. The first level, dot-blot hybridization is capable of dividing the organisms into their respective clades. Real time PCR is the second level and able is to detect specific groups within each clade. Microarrays take it a step further and are capable of identifying even more organisms within each clade. Although a limited number of samples were analyzed by microarrays, this technology shows promise and this study was viewed as a pilot for their application. The *rbcL* probes designed were based upon published sequences from 2003 and we now have a much greater understanding of the diversity of *rbcL*-containing phytoplanktonic phylotypes. Future studies should employ this knowledge for judicious probe selection.

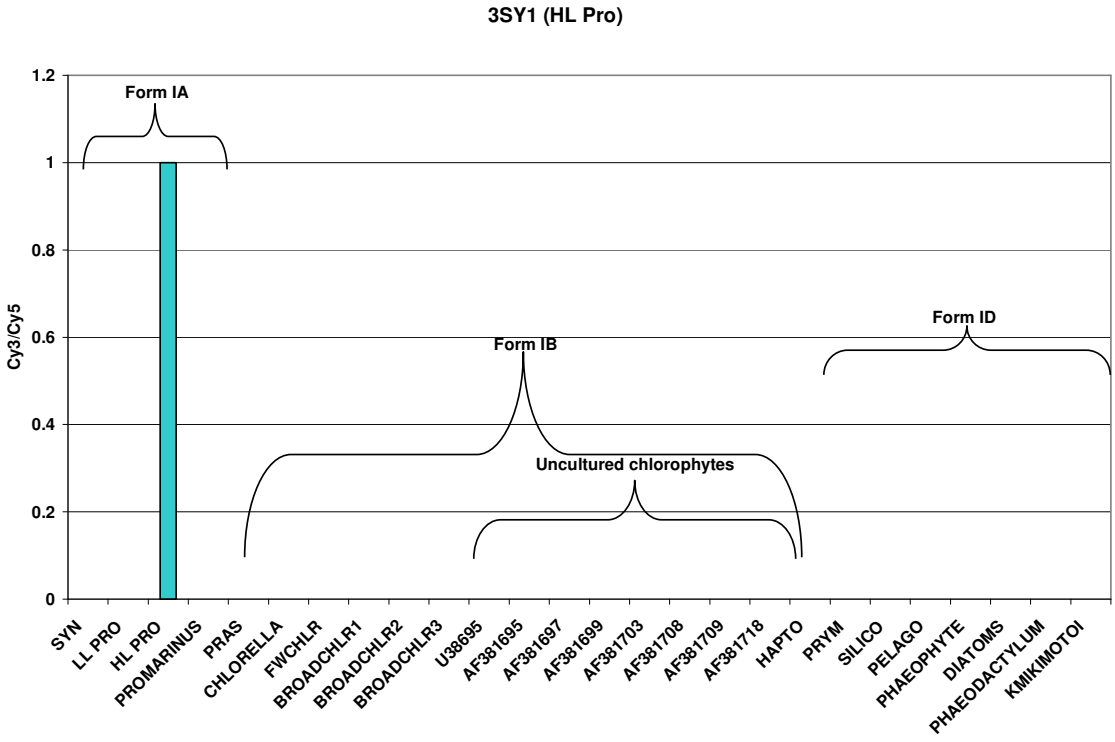


Figure 3.3. Hybridization of 3SY1 (HL *Prochlorococcus*) to array.

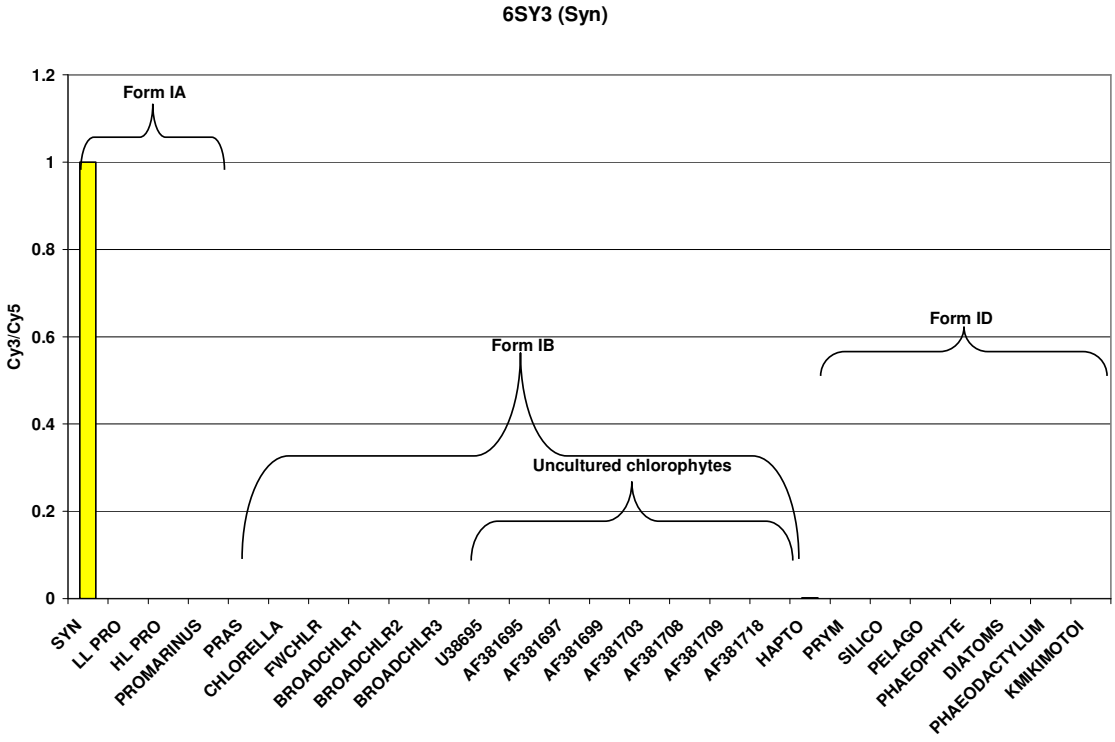


Figure 3.4. Hybridization of 6SY3 (*Synechococcus*) to array.

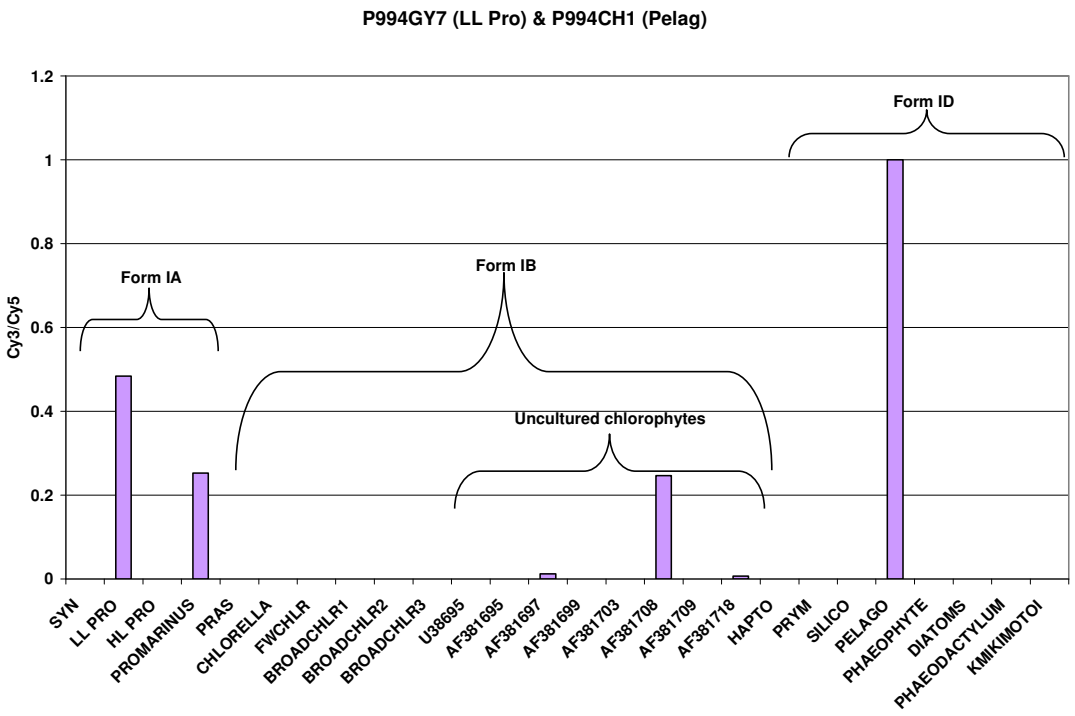


Figure 3.5. Hybridization of P994GY7 (LL *Prochlorococcus*) and P994CH1 (pelagophyte) to array.

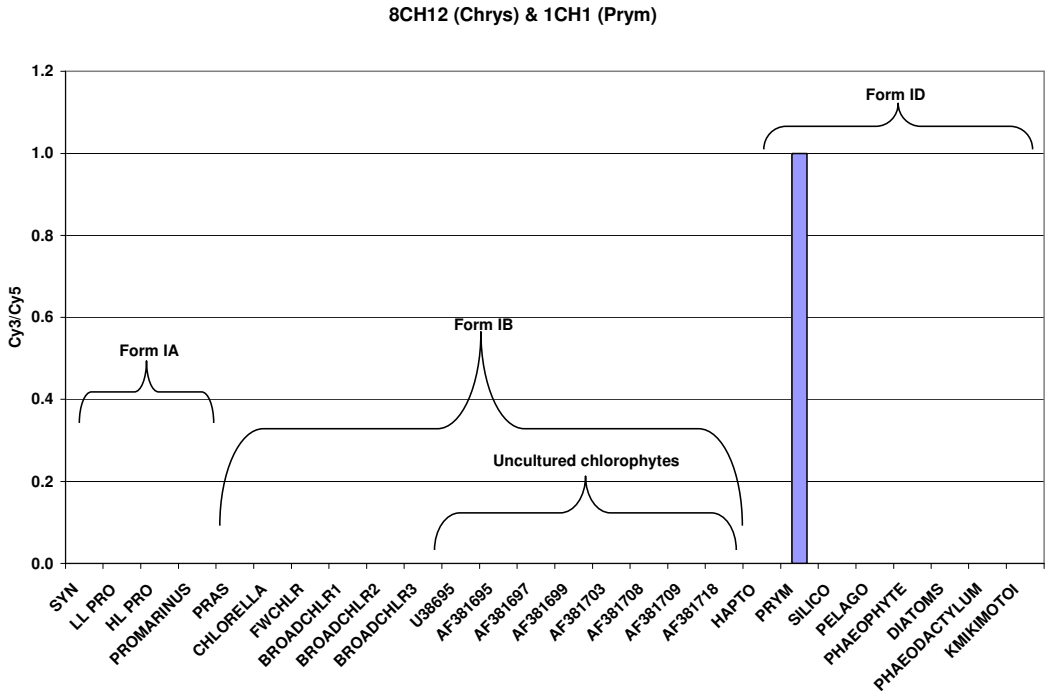


Figure 3.6. Hybridization of 8CH12 (chrysophyte) and 1CH1 (prymnesiophyte) to array.

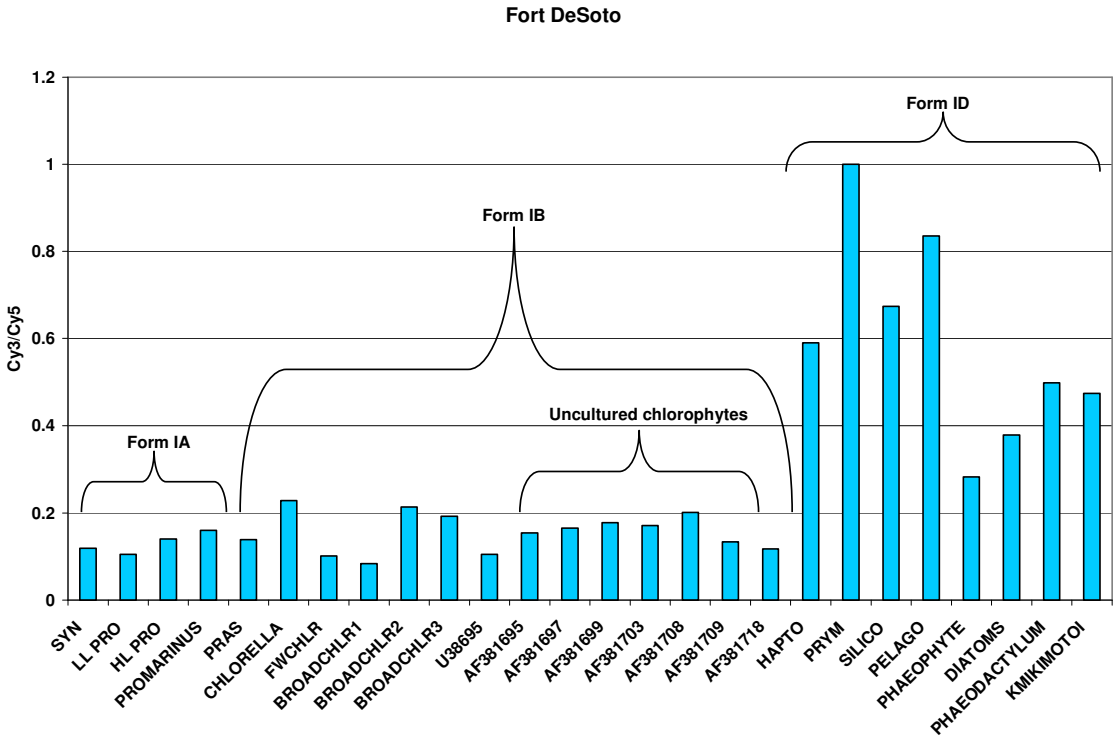


Figure 3.7. Hybridization of cDNA made from water collected from Fort Desoto and amplified with Form ID primers.

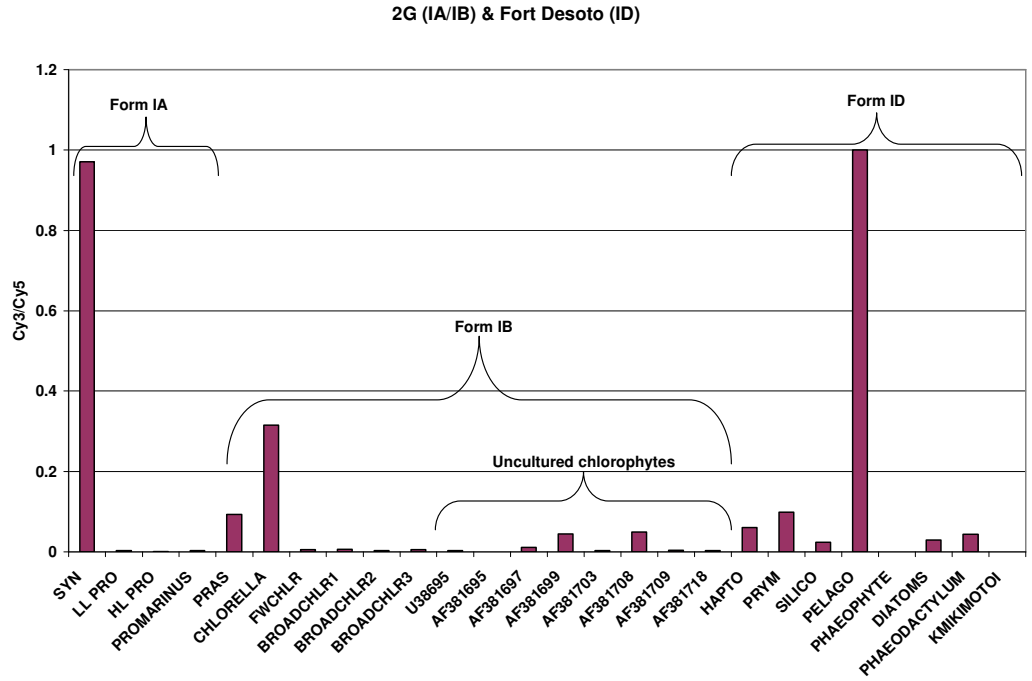


Figure 3.8. Hybridization of cDNA made from water collected from Fort Desoto and Station 2G. cDNA obtained from Station 2G was amplified with Form IA/IB primers while cDNA from Fort Desoto was amplified with Form ID primers.

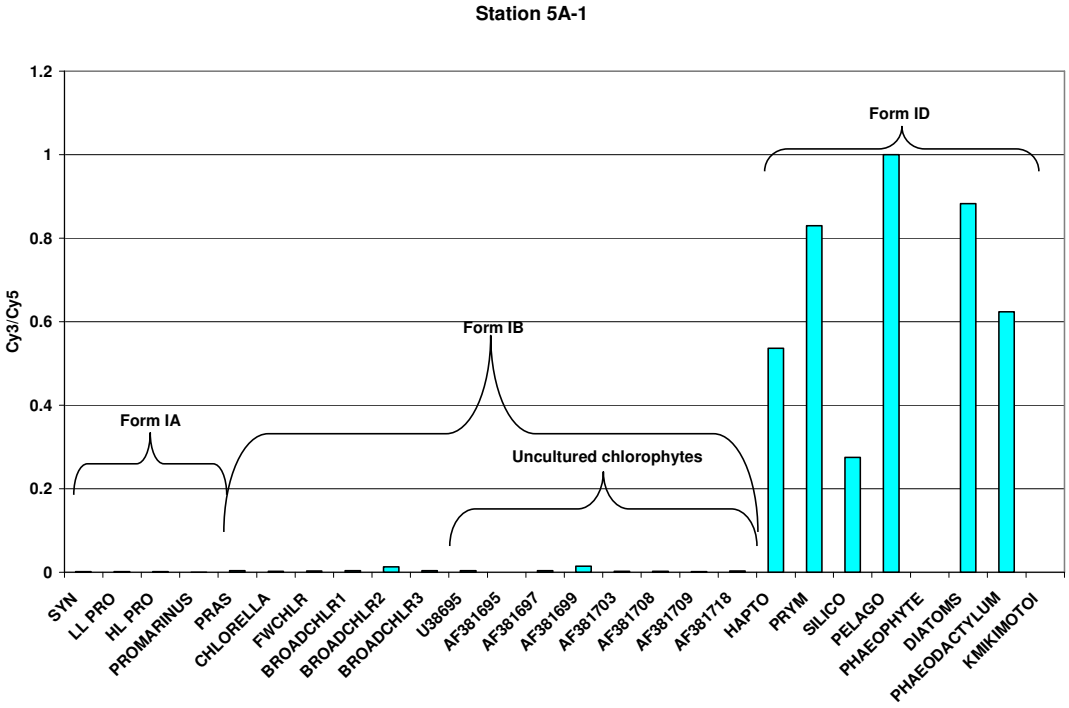


Figure 3.9. Hybridization of cDNA obtained from Station 5A. cDNA used in this experiment was amplified with Form ID primers.

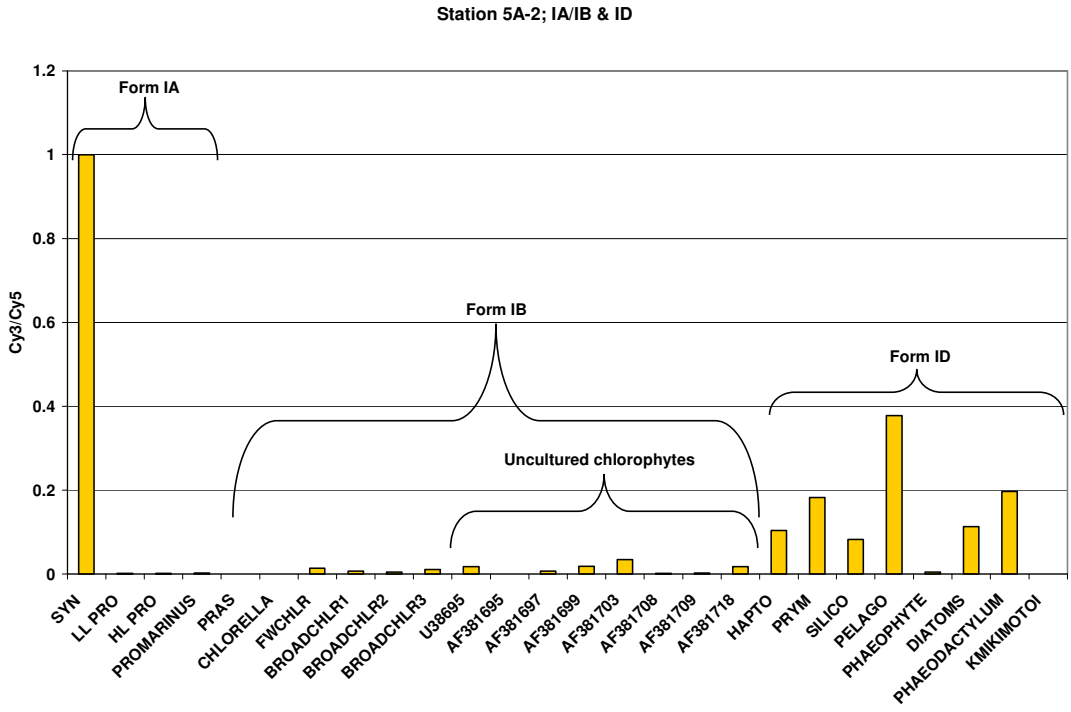


Figure 3.10. Hybridization of cDNA obtained from Station 5A. cDNA used in this experiment was amplified with Form IA/IB and Form ID primers.

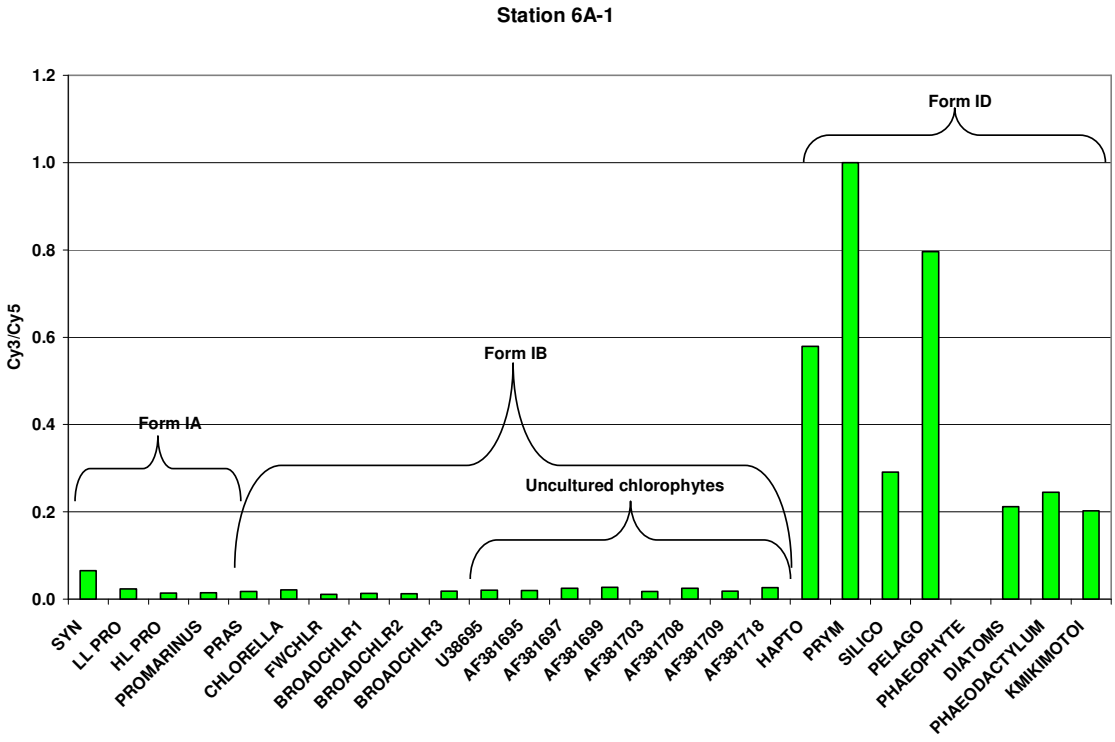


Figure 3.11. Hybridization of cDNA obtained from Station 6A that was amplified with Form ID primers. This cDNA was sent to the Ward Lab at Princeton University in July 2005.

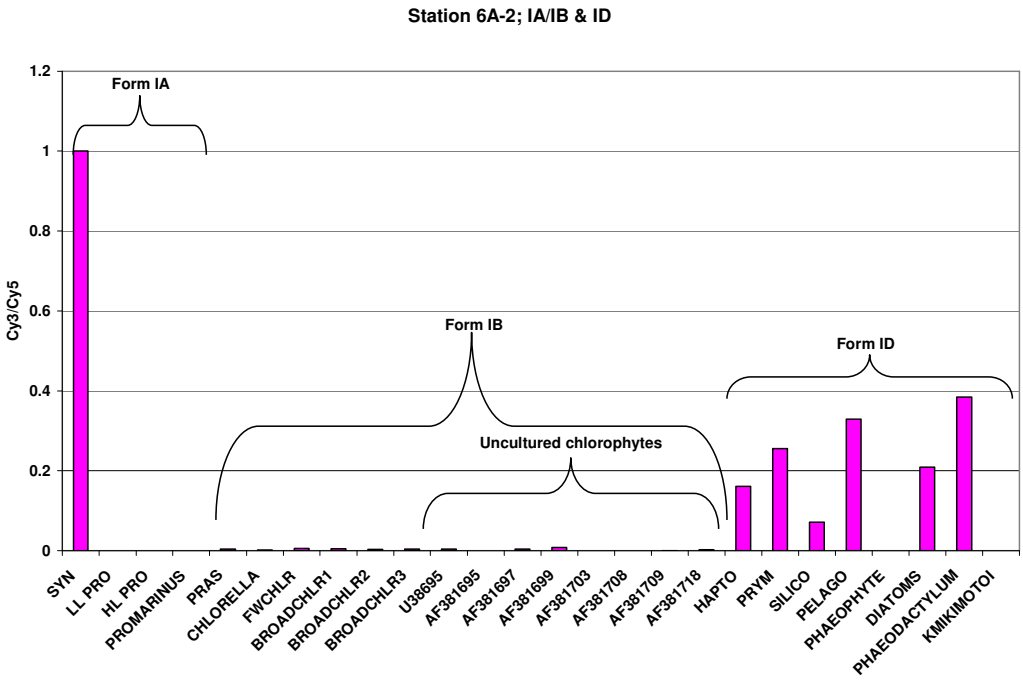


Figure 3.12. Hybridization of cDNA obtained form Station 6A that was amplified with both Form IA/IB and Form ID primers. cDNA was sent to the Ward Lab at Princeton University in November 2005.

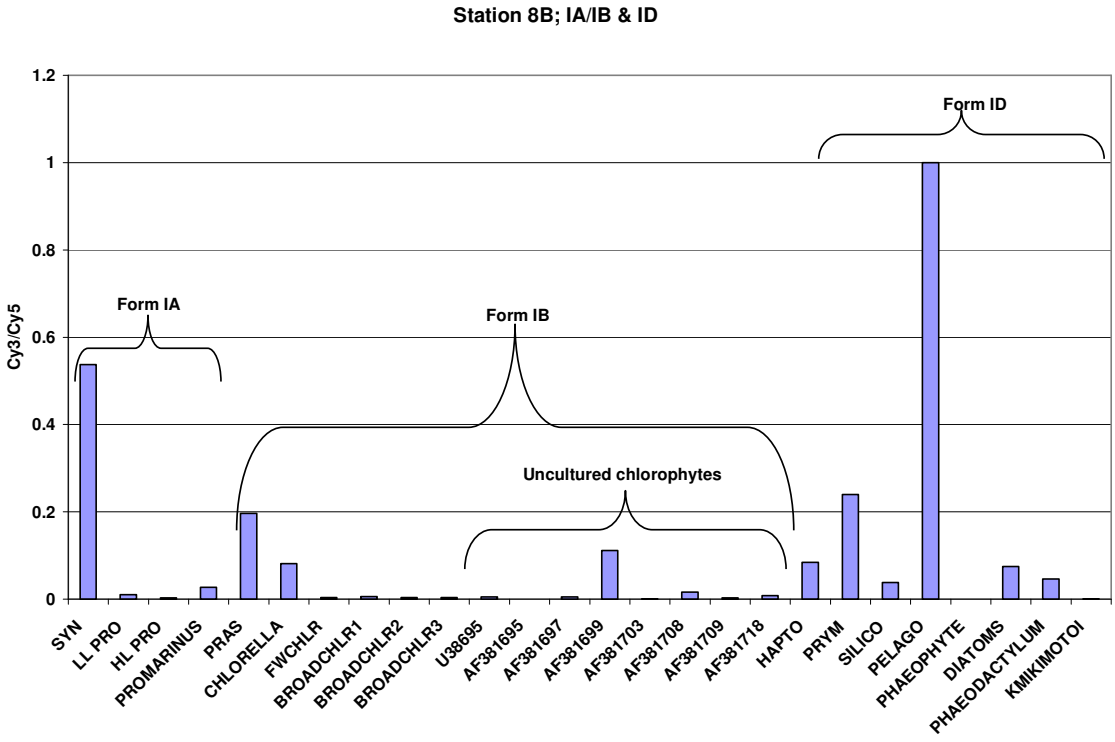


Figure 3.13. Hybridization of cDNA obtained from Station 8B. cDNA was amplified with Form IA/IB and Form ID primers.

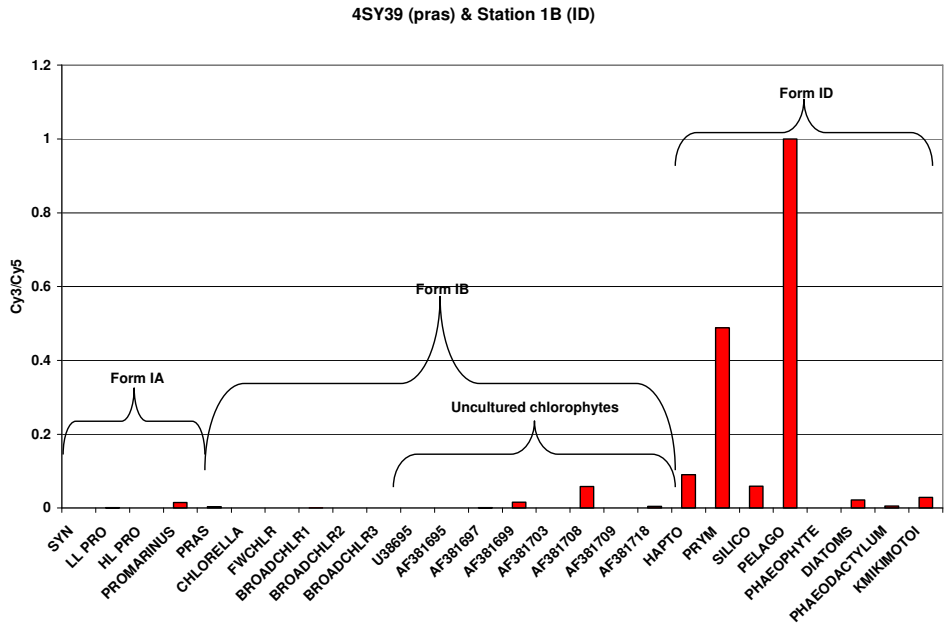


Figure 3.14. Hybridization of cDNA obtained from Station 1B and transcript 4SY39 (a prasinophyte). cDNA from Station 1B was amplified with Form ID primers.

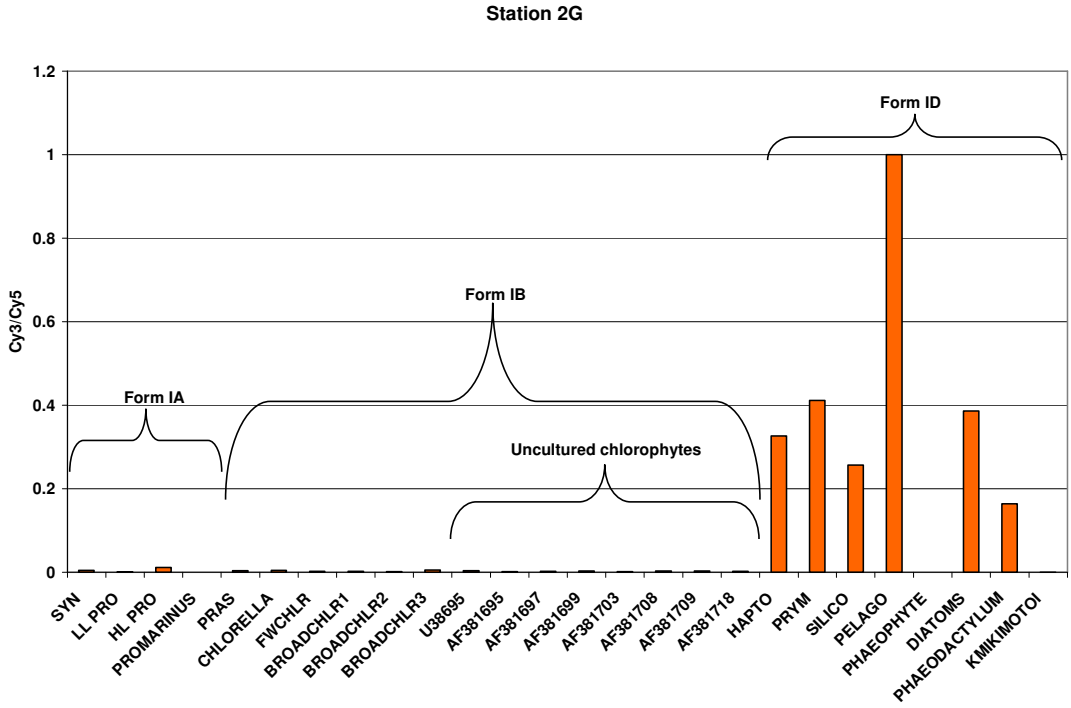


Figure 3.15. Hybridization of cDNA obtained from Station 2G. cDNA was amplified with Form ID primers.

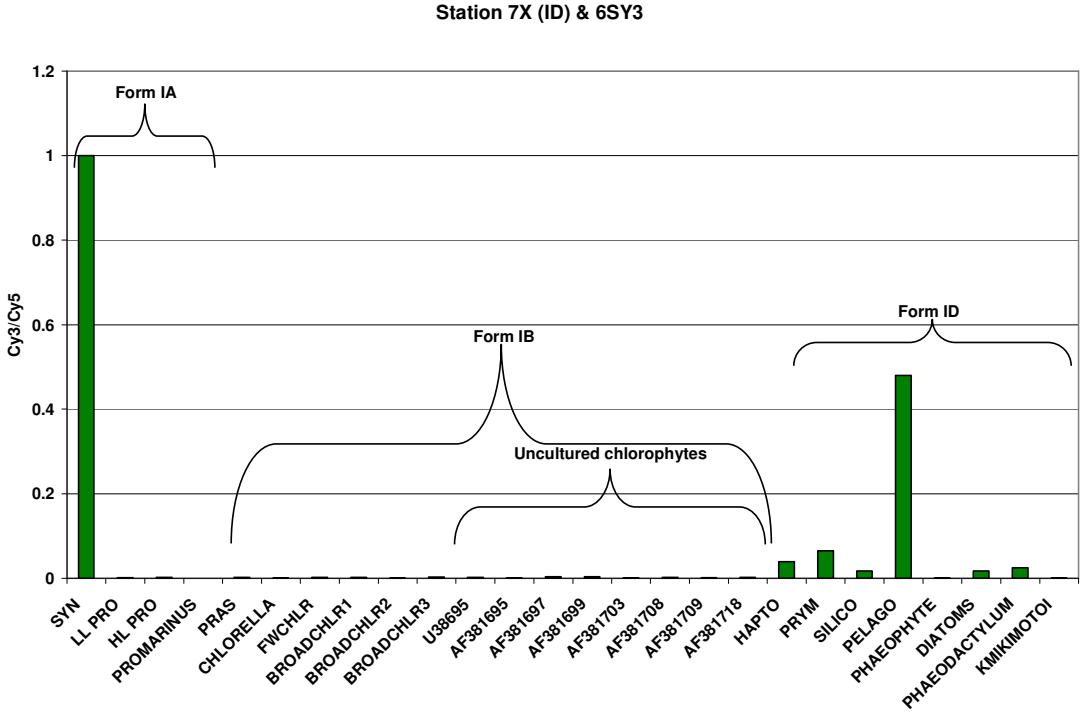


Figure 3.16 . Hybridization of cDNA obtained from Station 7X and clone 6SY3 (a synechococcus). cDNA from Staion 7X was amplified with Form ID primers.

List of References

1. **Amon, R. M. W., and R. Benner.** 1998. Seasonal Patterns of Bacterial Abundance and Production in the Mississippi River Plume and Their Importance for the Fate of Enhanced Primary Production. Microb. Ecol. **35:**289-300.

2. **Atomi, H.** 2002. Microbial enzymes involved in carbon dioxide fixation. J. Biosci. Bioeng. **94:**497-505.

3. **Badger, M. R., and G. D. Price.** 1994. The Role of Carbonic Anhydrase in Photosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. **45:**369-392.

4. **Badger, M. R., G. D. Price, B. M. Long, and F. J. Woodger.** 2006. The environmental plasticity and ecological genomics of the cyanobacterial CO2 concentrating mechanism. J. Exp. Bot. **57:**249-265.

5. **Binder, B. J., S. W. Chisholm, R. J. Olson, S. L. Frankel, and A. Z. Worden.** 1996. Dynamics of picophytoplankton, ultraphytoplankton and bacteria in the Central Equatorial Pacific. Deep-Sea Research(Part II, Topical Studies in Oceanography) **43:**907-931.

6. **Blanchot, J., and M. Rodier.** 1996. Picophytoplankton abundance and biomass in the western tropical Pacific Ocean during the 1992 El Nino year: results from flow cytometry. Deep-Sea Res. Part I **43:**877-895.

7. **Chen, Y. B., B. Dominic, M. T. Mellon, and J. P. Zehr.** 1998. Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium Trichodesmium sp. strain IMS 101. J. Bacteriol. **180:**3598-3605.

8. **Chung, C. C., S. P. Hwang, and J. Chang.** 2003. Identification of a high-affinity phosphate transporter gene in a prasinophyte alga, Tetraselmis chui, and its expression under nutrient limitation. Appl. Environ. Microbiol. **69:**754-759.

9. **Corredor, J. E., B. Wawrik, J. H. Paul, H. Tran, L. Kerkhof, J. M. López, A. Dieppa, and O. Cárdenas.** 2004. Geochemical Rate-RNA Integration Study: Ribulose-1, 5-Bisphosphate Carboxylase/Oxygenase Gene Transcription and Photosynthetic Capacity of Planktonic Photoautotrophs. Appl. Environ. Microbiol. **70:**5459-5468.

10. **del Giorgio, P. A., and C. M. Duarte.** 2002. Respiration in the open ocean. Nature **420:**379-384.

11. **Ellis, R. J.** 1979. The most abundant protein in the world. Trends Biochem. Sci. **4:**1—244.

12. **Elsaied, H., and T. Naganuma.** 2001. Phylogenetic Diversity of Ribulose-1, 5-Bisphosphate Carboxylase/Oxygenase Large-Subunit Genes from Deep-Sea Microorganisms. Appl. Environ. Microbiol. **67:**1751-1765.

13. **Fukuzawa, H., K. Miura, K. Ishizaki, K. Kucho, T. Saito, T. Kohinata, and K. Ohyama.** 2001. Ccm1, a regulatory gene controlling the induction of a carbon-concentrating mechanism in Chlamydomonas reinhardtii by sensing CO2 availability. PNAS **98:**5347.

14. **Giordano, M., J. Beardall, and J. A. Raven.** 2005. CO2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. Annu. Rev. Plant. Biol. **56:**99-131.

15. **Hartman, F. C., and M. R. Harpel.** 1994. Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. Annu. Rev. Biochem. **63:**197-234.

16. **Holland, P. M., R. D. Abramson, R. Watson, and D. H. Gelfand.** 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:**7276-7280.

17. **Jenkins, B. D., G. F. Steward, S. M. Short, B. B. Ward, and J. P. Zehr.** 2004. Fingerprinting Diazotroph Communities in the Chesapeake Bay by Using a DNA Macroarray. Appl. Environ. Microbiol. **70:**1767-1776.

18. **John, D. E., Z. A. Wang, X. Liu, R. H. Byrne, J. E. Corredor, J. M. Lopez, A. Cabrera, D. A. Bronk, and J. H. Paul.** Submitted. Carbon fixation gene (RuBisCO) transcripts and CO2 flux in the Mississippi River Plume. Int. Soc. Microb. Ecol. .

19. **Kaplan, A., and L. Reinhold.** 1999. CO2 CONCENTRATING MECHANISMS IN PHOTOSYNTHETIC MICROORGANISMS. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50:**539-570.

20. **Letowski, J. G., R. G. Brousseau, and L. G. Masson.** 2003. DNA Microarray Applications in Environmental Microbiology. Anal. Lett. **36:**3165-3184.

21. **Loy, A., C. Schulz, S. Lücker, A. Schöpfer-Wendels, K. Stoecker, C. Baranyi, A. Lehner, and M. Wagner.** 2005. 16S rRNA Gene-Based Oligonucleotide

Microarray for Environmental Monitoring of the Betaproteobacterial Order" Rhodocyclales. Appl. Environ. Microbiol. **71:**1373-1386.

22. **Marcus, Y., J. A. Berry, and J. Pierce.** 1992. Photosynthesis and photorespiration in a mutant of the cyanobacterium Synechocystis PCC 6803 lacking carboxysomes. Planta **187:**511-516.

23. **McGinn, P. J., G. D. Price, R. Maleszka, and M. R. Badger.** 2003. Inorganic Carbon Limitation and Light Control the Expression of Transcripts Related to the CO2-Concentrating Mechanism in the Cyanobacterium Synechocystis sp. Strain PCC6803. Plant Physiol. **132:**218-229.

24. **Miller, C. B.** 2004. Biological Oceanography. Blackwell Pub Malden, MA.

25. **Miziorko, H. M., and G. H. Lorimer.** 1983. Ribulose-1,5-bisphosphate carboxylase-oxygenase. Annu. Rev. Biochem. **52:**507-535.

26. **Morse, D., P. Salois, P. Markovic, and J. W. Hastings.** 1995. A nuclear-encoded form II RuBisCO in dinoflagellates. Science **268:**1622-1624.

27. **Nelson, N., and A. Ben-Shem.** 2004. The complex architecture of oxygenic photosynthesis. Nat. Rev. Mol. Cell Biol. **5:**971-982.

28. **Paul, J. H., L. Cazares, and J. Thurmond.** 1990. Amplification of the rbcL gene from dissolved and particulate DNA from aquatic environments. Appl. Environ. Microbiol. **56:**1963-1966.

29. **Paul, J. H., J. B. Kang, and F. R. Tabita.** 2000. Diel Patterns of Regulation of rbcL Transcription in a Cyanobacterium and a Prymnesiophyte. Mar. Biotechnol. **2:**429-436.

30. **Paul, J. H., and B. Myers.** 1982. Fluorometric Determination of DNA in Aquatic Microorganisms by Use of Hoechst 33258. Appl. Environ. Microbiol. **43:**1393-1399.

31. **Raven, J. A.** 1996. Inorganic carbon assimilation by marine biota. J. Exp. Mar. Biol. Ecol. **203:**39-47.

32. **Redalje, D. G., S. E. Lohrenz, and G. L. Fahnenstiel.** 1994. The Relationship between Primary Production and the Vertical Export of Particulate Organic Matter in a River-Impacted Coastal Ecosystem. Estuaries **17:**829-838.

33. **Roszak, D. B., Colwell, R. R.** 1987. Survival strategies of bacteria in the natural environment. Microbiology and Molecular Biology Reviews **51:**365-379.

34. **Scharf, M. E., D. Wu-Scharf, B. R. Pittendrigh, and G. W. Bennett.** 2003. Caste- and development-associated gene expression in a lower termite. Genome Biol. **4:**R62-R62.11.

35. **Schneider, G., Y. Lindqvist, and C. I. Branden.** 1992. RUBISCO: structure and mechanism. Annu. Rev. Biophys. Biomol. Struct. **21:**119-143.

36. **So, A. K. C., G. S. Espie, E. B. Williams, J. M. Shively, S. Heinhorst, and G. C. Cannon.** 2004. A Novel Evolutionary Lineage of Carbonic Anhydrase (ϵ Class) Is a Component of the Carboxysome Shell. J. Bacteriol. **186:**623-630.

37. **Spalding, M. H., K. Van, Y. Wang, and Y. Nakamura.** 2002. Acclimation of Chlamydomonas to changing carbon availability. Functional Plant Biology **29:**221-230.

38. **Steward, G. F., B. D. Jenkins, B. B. Ward, and J. P. Zehr.** 2004. Development and Testing of a DNA Macroarray To Assess Nitrogenase (nifH) Gene Diversity. Appl. Environ. Microbiol. **70:**1455-1465.

39. **Tabita, F. R. M.** 1999. Microbial ribulose 1, 5-bisphosphate carboxylase/oxygenase: A different perspective. Photosynthesis Res. **60:**1-28.

40. **Tabita, F. R.** 1988. Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms. Microbiol. Rev. **52:**155-189.

41. **Taroncher-Oldenburg, G., E. M. Griner, C. A. Francis, and B. B. Ward.** 2003. Oligonucleotide Microarray for the Study of Functional Gene Diversity in the Nitrogen Cycle in the Environment. Appl. Environ. Microbiol. **69:**1159-1171.

42. **Turner, R. E., and N. N. Rabalais.** 1991. Changes in Mississippi River Water Quality This Century. Bioscience **41:**140-147.

43. **U.S. Environmental Protection Agency.** 2007. Mississippi River Basin and Gulf of Mexico hypoxia: culture and history. [online]: http://www.epa.gov/mbasin/culture.htm

44. **Ward, B. B.** 2005. MOLECULAR APPROACHES TO MARINE MICROBIAL ECOLOGY AND THE MARINE NITROGEN CYCLE. Annu. Rev. Earth Planet. Sci. **33:**301-333.

45. **Watson, G. M., and F. R. Tabita.** 1997. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a molecule for phylogenetic and enzymological investigation. FEMS Microbiol. Lett. **146:**13-22.

46. **Wawrik, B., J. H. Paul, L. Campbell, D. Griffin, L. Houchin, A. Fuentes-Ortega, and F. Muller-Karger.** 2003. Vertical structure of the phytoplankton community

associated with a coastal plume in the Gulf of Mexico. Mar. Ecol. Prog. Ser. **251:**87–101.

47. **Wawrik, B., and J. H. Paul.** 2004. Phytoplankton community structure and productivity along the axis of the Mississippi River plume in oligotrophic Gulf of Mexico waters. Aquat. Microb. Ecol. **35:**185-196.

48. **Wawrik, B., J. H. Paul, and F. R. Tabita.** 2002. Real-time PCR quantification of rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes. Appl. Environ. Microbiol. **68:**3771-3779.

49. **Wikipedia.** 2007. RuBisCO. [online]: http://en.wikipedia.org/wiki/RuBisCO

50. **Wilkins, M. B.** 1992. Circadian rhythms: their origin and control. New Phytol. **121:**347-375.

51. **Woodger, F. J., M. R. Badger, and G. D. Price.** 2003. Inorganic Carbon Limitation Induces Transcripts Encoding Components of the CO2-Concentrating Mechanism in Synechococcus sp. PCC7942 through a Redox-Independent Pathway. Plant Physiol. **133:**2069-2080.

52. **Wyman, M., J. T. Davies, S. Hodgson, G. A. Tarran, and D. A. Purdie.** 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 {dagger}. Appl. Environ. Microbiol. **71:**1654-1661.

Appendices

Appendix A

Table A-1. *rbcL* probe list

rbcL Probe Sequences Probe Name	Component sequences Accession Number	Identity	Probe Sequence
K. mikimotoi	AB034635	Karenia mikimotoi	GATCCCCGGGAATTGCCATGCCTCTAATGATTAAAGGTTTCTACGACACTCTGCTTGATTCTAAGATTNGAAATTGACCTACCTCAAGGTA
P. tricornutum	AF195952	Phaeodactylum tricornutum	GATCCCCGGGAATTGCCATGCCTTTAATGATTAAAGGTTTCTACGATACTTTATTACTAACTCATTTAAATGTTAATTTACCATACGGTA
Diatoms	consensus		GATCCCCGGGAATTGCCATGCCTTTAATGATTAAAGGTTTCTACGATATTTTACGTTTAACAAAATTAGATGTTAACTTACCATTCGGTA
	AB018006	Detonula confervacea	
	AB018007	Thalassiosira nordenskioldii	
	AF015568	Rhizosolenia setigera	
	AF015569	Skeletonema costatum	
	AF064744	Mischococcus sphaerocephalus	
	AF069499	Pleurochloridella botrydiopsis	
	AF109210	Thalassiosira pseudonana	
	AF179005	Bacillariophyta gen. Sp.	
	AF179008	Bacillariophyta gen. Sp.	
	AF195952	Phaeodactylum tricornutum	
	AF333977	Bolidomonas mediterranea	
	AF333978	Bolidomonas pacifica	
	AF333979	Bolidomonas pacifica	
	AF372696	Bolidomonas pacifica strain CCMP 1866	
	AF381678	uncultured stramenophile clone	
	AF381682	uncultured stramenophile clone	
	AY157391	uncultured marine euk	
	AY157396	uncultured marine euk	
	AY157398	uncultured marine euk	
	AY157404-AY157440	uncultured marine euk	
	AY278368-AY278392	uncultured marine euk	
	AY356332	uncultured marine euk	
	M59080	Cylindrotheca sp.	
	UDU31876	unidentified diatom endosymbiont of Peridinium	
	U43061-U04373	unidentified algae	
	AF381676	Uncultured stramenophile clone 994CH22	
	AF381681	Uncultured stramenophile clone 994FH13	
	AF381684	Uncultured stramenophile clone 994HH14	
Phaeophytes	consensus		GATCCCCGGGAATTGCCATGCCTTTAATGGTCAGAGGATTCTACAATACGTTATTATTAACTCAGTTAAAAATTAATTTAGCTGAAGGTT
	AB045259	Phaeiphonielia cryophila	
	AF055400	Haplospora globosa	
Pelagophytes	consensus		GATCCCCGGGAATTGCCATGCCATTAATGATCCAAGGTTTCTACGACACATTATTAAAGACTAAGTTATCTATCGATTTACCTAAGGGTA
	U89698	Pelagomonas calceolata	
	AF117904	Aureococcus anophagefferens	
	AF117905	Aureococcus anophagefferens	
	AF117906	Aureococcus anophagefferens	
	AF381685	Uncultured stramenophile 994CH1	
	AF381686	Uncultured stramenophile clone 994CH2	
	AF381691	Uncultured stramenophile clone 994HH7	
Silicoflagellates	AF381735	Uncultured haptophyte cone 994AH8	GATCCCCGGGAATTGCCATGCCTCTAATGGTTAAAGGTTTCTACAACACTTTATTAAACGTTAAGACATCTGTAAACCTTCCTCAAGGTT
Prymnesiophytes	AF381751	Uncultured haptophyte clone 994GH21	GATCCCCGGGAATTGCCATGCCTCTAATGATTAAAGGTTTCTACAACACTCTTCTAGATTTTAAATCTGACATCAACTTACCCTGTGGTA
Haptoconsensus	consensus		GATCCCCGGGAATTGCCATGCCTCTAATGATTAAAGGTTTCTACAACACTCTTCTTGATTTTAAGACTGATGTTAACCTTCCTCAAGGTT
	AF381741	Uncultured haptophyte clone 994CH9	
	AF381747	Uncultured haptophyte clone 994EH20	
	D45844	Gephyrocapsa oceanica	
	D45845	Emiliania hyxleyi	
	AB043693	Isochrysis galbana	
	AB043695	Chrysochromulina alifera	
	AB043700	Chrysochromulina spinifera	
	D45847	Pavlova salina	
	AF381733	Uncultured haptophyte clone 994AH5	
	AF381738	Uncultured haptophyte clone 994BH3	
	AB043689	Cruciplacolithus neohelis	
	AF381748	Uncultured haptophyte clone 994EH4	
	AF381742	Uncultured haptophyte clone 994DH1	
	AF381734	Uncultured haptophyte clone 994AH7	
	AF381736	Uncultured haptophyte clone 994BH15	
	AB043694	Chrysochromulina parva	

Appendix A (Continued)
Table A-1 (Continued)

	AB043696	Imantonia rotunda	
	AB043697	Chrysochromulina sp. TKB8936	
	AB043699	Platychrysis sp. TKB8934	
	AF381749	Uncultured haptophyte 994EH9	
	AF381732	Uncultured haptophyte clone 994AH28	
	D45846	Chrysochromulina hirta	
	AF381740	Uncultured haptophyte 994CH25	
	AF381745	Uncultured haptophyte clone 994EH1	
	AF381739	Uncultured haptophyte clone 994BH5	
	AF381746	Uncultured haptophye clone 994EH2	
	AF381752	Uncultured haptophyte clone 994HH11	
	AB043688	Pleurochrysis haptonemofera	
	AB043691	Umbilicosphaera sibogae	
	AB043690	Calcidiscus leptoporus	
	AB043692	Hellcosphaera carteri	
	D45842	Calyptrosphaera sphaeroidea	
	AF381743	Uncultured haptophyte clone 994DH2	
	AF381737	Uncultured haptophyte clone 994BH22	
	AB043689	Cruciplacolithus neohelis	
Chlorophytes			
ChlGrp1/Pras	consensus		GATCCCCGGGAATTGCCATGCGTAACGTAACGCTTGGTTTCGTAGACATTATGCGTGATGCATACATTGAGAAGGACCGTGACCGTGGAG
	AF381706	Uncultured chlorophyte clone 994EY3	
	AF381707	Uncultured chlorophyte clone 994EY4	
	AF381711	Uncultured chlorophyte clone 994FY19	
	AF381712	Uncultured chlorophyte clone 994FY23	
	AF381713	Uncultured chlorophyte clone 994FY27	
	AF381715	Uncultured chlorophyte clone 994FY3	
ChlrGrp2/Chlorella	consensus		GATCCCCGGGAATTGCCATGCGTGAAGTAACTTTAGGTTTCGTAGACCTAATGCGTGATGCTTACATCGAGAAGGATCGTTCACGTGGTG
	AF381701	Uncultured chlorophyte clone 99CY2	
	AF381704	Uncultured chlorophyte clone 994DY5F	
	AF381705	Uncultured chlorophyte clone 994EY1	
	AF381714	Uncultured chlorophyte clone 994FY29	
	AF381716	Uncultured chlorphyte clone 994FY43	
	AF381717	Uncultured chlorophyte clone 994FY8	
ChlrGrp3/FWChlr	consensus		GATCCCCGGGAATTGCCATGCGTATTTCTCCTGCTTACGTTAAAACATTCCAAGGTCCACCACACGGTATTCAGGTTGAACGTGACAAAT
	AB076098	Volvox tertius	
	AB076099	Volvox carteri f. nagariensis	
	AB076100	Volvox carteri f. weismannia	
	AB084332	Scenedesmus quadricauda	
	AB084335	Pseudocarteria mucosa	
	AB086174	Volvox tertius	
	U80809	Chloromonas sp. ANT3	
ChlrGrp4/BroadChlr1	consensus		GATCCCCGGGAATTGCCATGCGTAATATTACTCTTGGTTTCGTTGATTTAATGCGTGATAACTATATTGCTAAGGATCGTGATCGTGGTG
	AF381693	Uncultured chlorophyte clone 994AY1	
	AF381696	Uncultured chlorophyte clone 994AY4	
ChlrGrp5/BroadChlr2	consensus		GATCCCCGGGAATTGCCATGAAAGATATTACTTTAGGTTTTGTTGATCTTATGAGAGATGACTTTGTAGAATTAGATCGATCTCGAGGAA
	AF381694	Uncultured chlorpphyte clone 994AY17	
	AF381698	Uncultured chlorophyte clone 994AY7	
ChlrGrp6/BroadChlr3	consensus		GATCCCCGGGAATTGCCATGAAAGCTTCGACCTTGGGCTTTGTTGACTTGATGCGCGAAGACCACATCGAAGCTGACCGCAGCCGTGGGG
	AF381702	Uncultured chlorophyte clone 994DY13F	
	AF381719	Uncultured chlorophyte clone 994GY	
	AF381720	Uncultured chlorophyte clone 994GY25	
single sequence probe	AF381695	Uncultured chlorophyte clone 994AY2	GATCCCCGGGAATTGCCATGAAAGATATTACTTTAGGATTTGTAGACTTGATGAGAGATGATTATGTTGAAGAAGATCGTCCTCGAGGTA
name = Acession number	AF391697	Uncultured chlorophyte clone 994AY6	GATCCCCGGGAATTGCCATGCGTGCTGTAACTTTAGGTTTCGTTGATTTAATGCGTGATAACGTTATTTTACGTGATAGATCGCGTGGTA
	AF381699	Uncultured chlorophyte clone 994AY6	GATCCCCGGGAATTGCCATGCGTGAGGTAACTCTTGGTTTCGTAGACCTTATGCGTGATGATTACGTAGAGAAGGACCGTAACCGTGGTA
	AF381703	Uncultured chlorophyte clone 994DY2F	GATCCCCGGGAATTGCCATGCGTGAAATTACTCTTGGTTTCGTCGATCTAATACGTGACAACCTAGTTGCAAAAGATCGCAGGCGTGGAA
	AF381708	Uncultured chlorophyte clone 994EY5	GATCCCCGGGAATTGCCATGCGTGAAGTTACTCTTGGTTTCGTTGATCTAATGCGTGATGCATATGTAGAAAAAGATCGTTCTCGTGGGA
	AF381709	Uncultured chlorophyte clone 994EY6	GATCCCCGGGAATTGCCATGCGCGAAATTACACTTGGCTTTGTCGACCTTATGCGCGATGACTTTATCTCTAAGGACCGTAGACGTGGAA
	AF381718	Uncultured chlorophyte clone 994GY1	GATCCCCGGGAATTGCCATGCGTCAAACAACTCTTGGCTATATCGATAACCTTCGTGAATCTTTTGTCCCTGAAGACCGTAGTCGCGGTA
	U38695	Cylindrocystis sp. UTEX1925	GATCCCCGGGAATTGCCATGCGTCAAGTGACTTTGGGTTTTGTAGACCTACTTCGTGATGATTACATTGAAAAAGATCGTAGTCGTGGTA
Prokaryotes			
ProkGrp1/Syn	consensus		GATCCCCGGGAATTGCCATGCGTCAGACCACCCTCGGCTACATCGACCAGCTGCGCGAATCCTTCGTGCCCGAAGACCGCAGCCGCGGCA
	AF381721	Uncultured Synechococcus sp. clone 99SY1	
	AF381722	Uncultured Synechococcus sp. clone 99SY12	
	AF381723	Uncultured Synechococcus sp. clone 99SY17	
ProkGrp2/LL Pro	consensus		GATCCCCGGGAATTGCCATGCGTCAGACCACXCTTGGTTATATTGACAACCTTCGTGAATCCTTTGTTCCTGAAGACCGCACACGCGGXA
	AF381660	Uncultured Proch sp. clone 994FY9F	
	AF381665	Uncultured Proch sp. clone 994GY7F	
	AF381666	Uncultured Proch sp. clone 994GY8F	
ProkGrp3/HL Pro	consensus		GATCCCCGGGAATTGCCATGCGTCAAACAACTCTTGGTTACATTGACAACTTAAGAGAGTCATTTGTTCCTGAAGATAGATCAAGAGGTA

Appendix A (Continued)

Table A-1 (Continued)

	AF381648	Uncultured Proch clone P99BY1	
	AF381649	Uncultured Proch clone 99BY1	
	AF381650	Uncultured Proch clone 99BY7	
	AF381651	Uncultured Proch 99CY1	
	AF381652	Uncultured Proch sp. clone 994DY10F	
	AF381653	Uncultured Proch sp. clone 994DY11F	
	AF381654	Uncultured Proch sp. clone 994DY16F	
ProkGrp4/ProMarinus	consensus		GATCCCCGGGAATTGCCATGCGTCAAACAACTCTTGGTTATATCGACAACCTACGTGAATCCTTTGTCCCTGAAGACCGTXCTCGCGGTA
	AY042086	Prochlorococcus marinus str. PAC1	
	AY042088	Prochlorococcus marinus subsp. marinus str. CCMP1375	
	AY042090	Prochlorococcus marinus subsp. Pastoris str. NATL1	

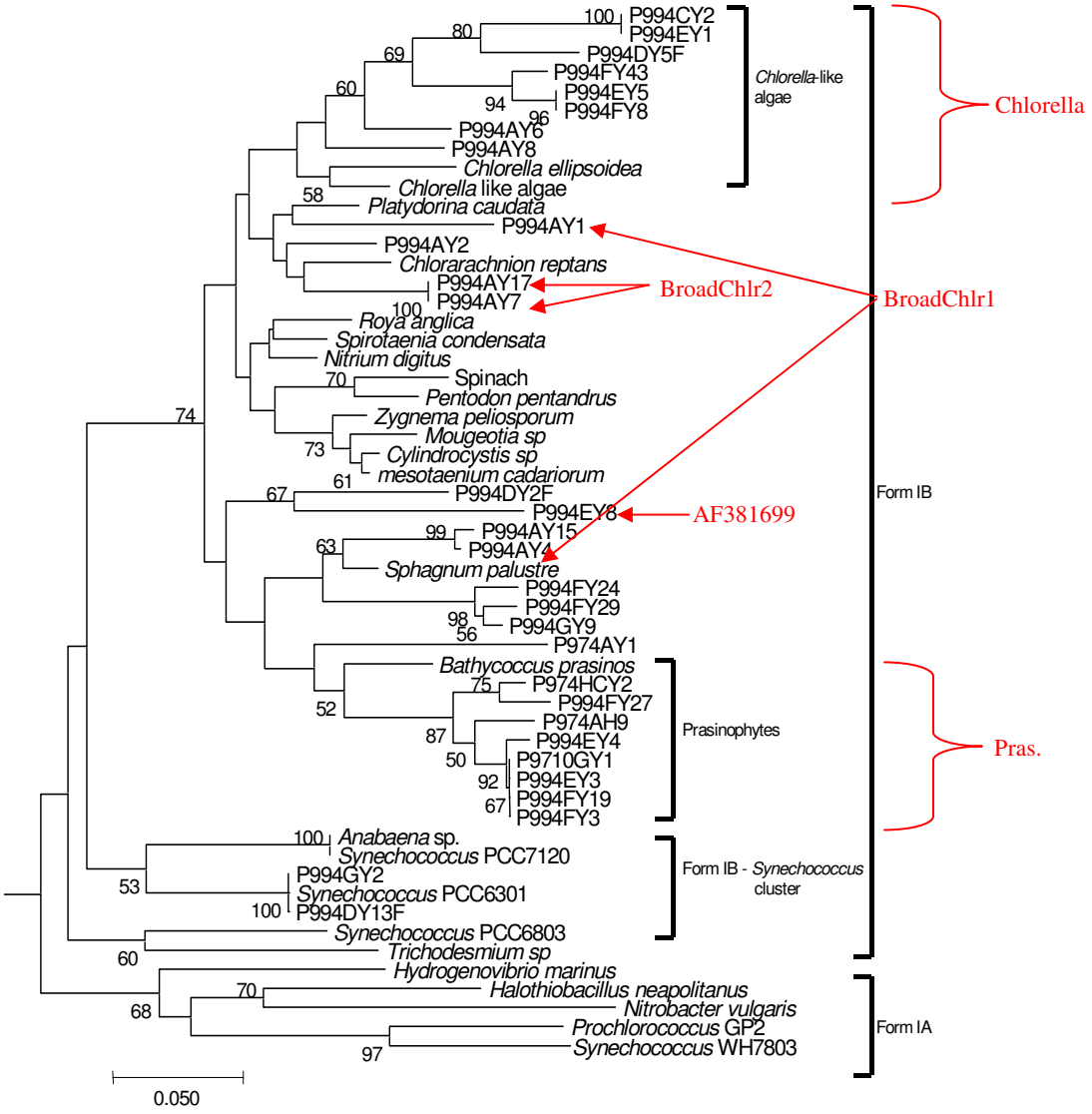


Figure A-1. Phylogenetic tree taken from Wawrik et al., 2003 showing the Form IB clones obtained from the Mississippi River Plume. The probe consensus groups present on the tree have been labeled.

Appendix A (Continued)

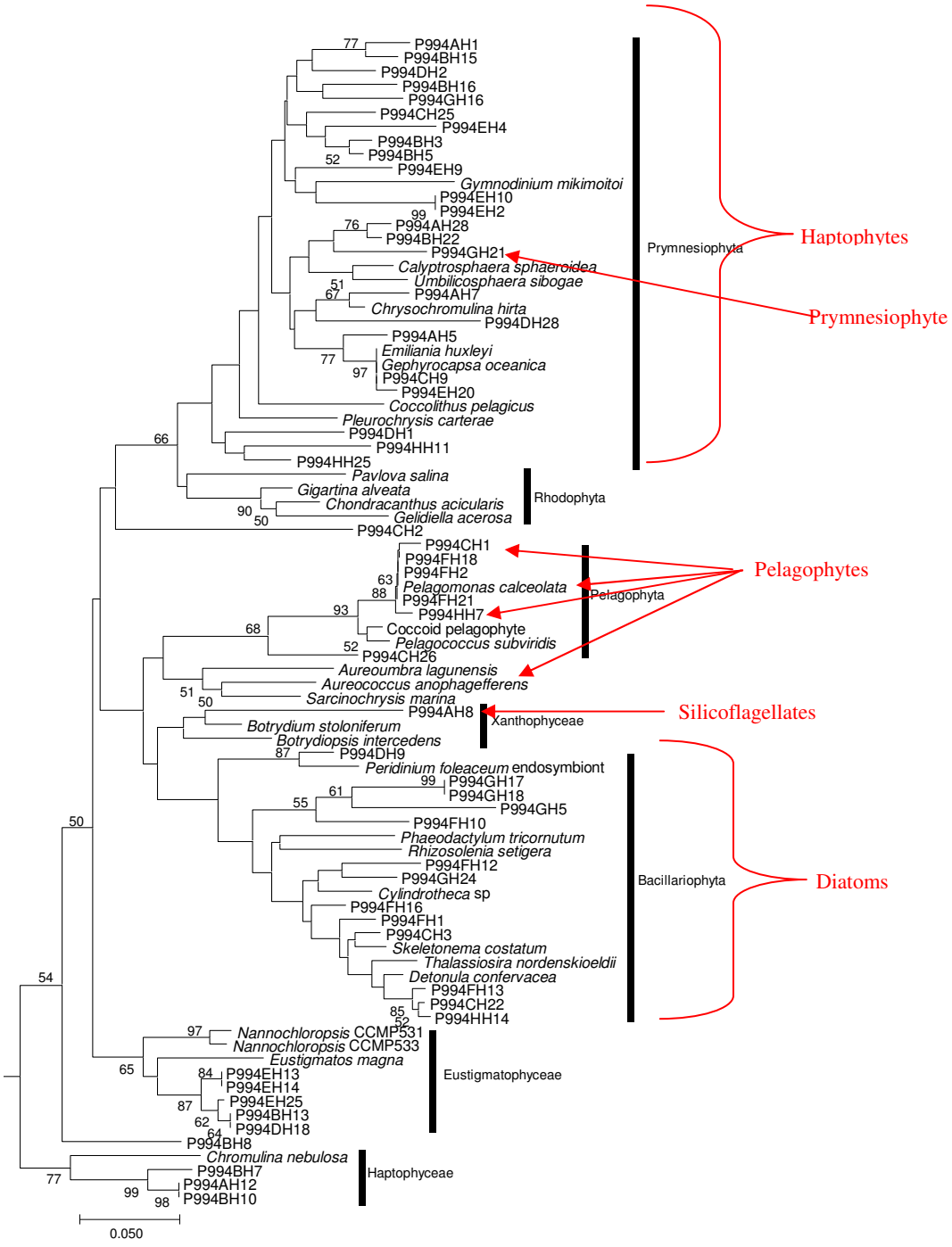


Figure A-2. Pylogenetic tree taken from Wawrik et al., 2003 of the Form ID sequences obtained from the Mississippi River Plume. The probe consensus groups present on the tree have been labeled.

Appendix A (Continued)

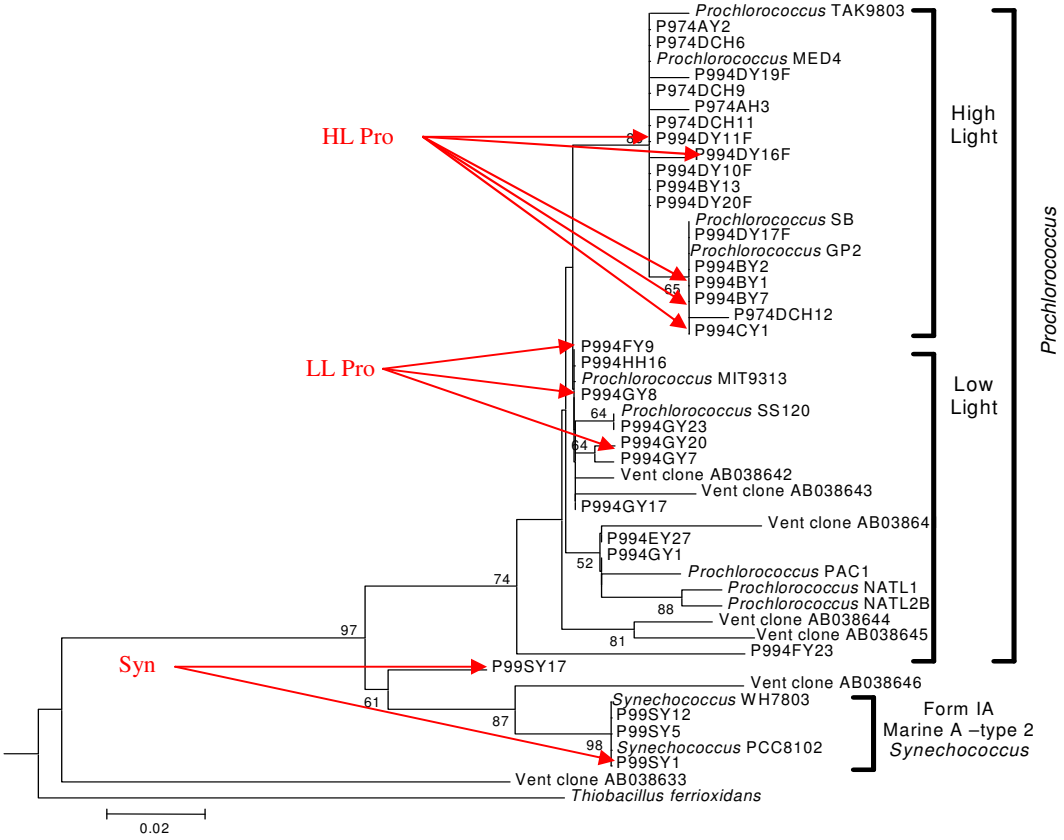


Figure A-3. Phylogenetic tree taken from Wawrik et al., 2003 showing the Form IA clones found in the Mississippi River Plume. The probe consensus groups present on the tree have been labeled

Appendix A (Continued)

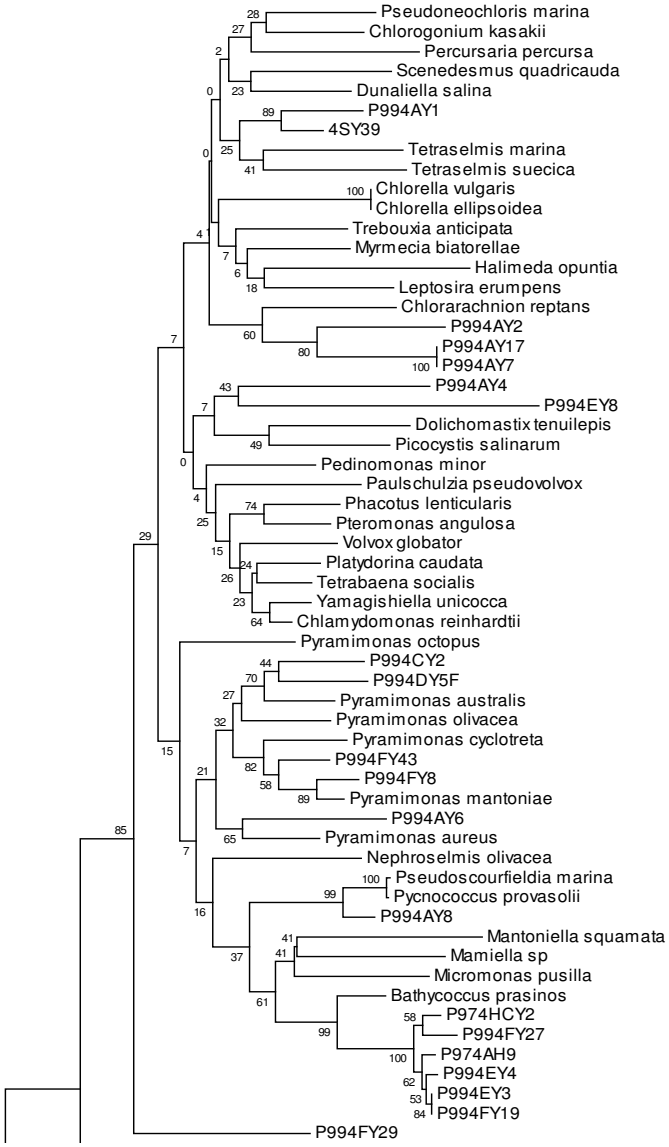


Figure A-4. Conensus tree of Form IB clone sequences obtained from the MRP in 1999 and 2001 along with their closest relatives.

Appendix A (Continued)

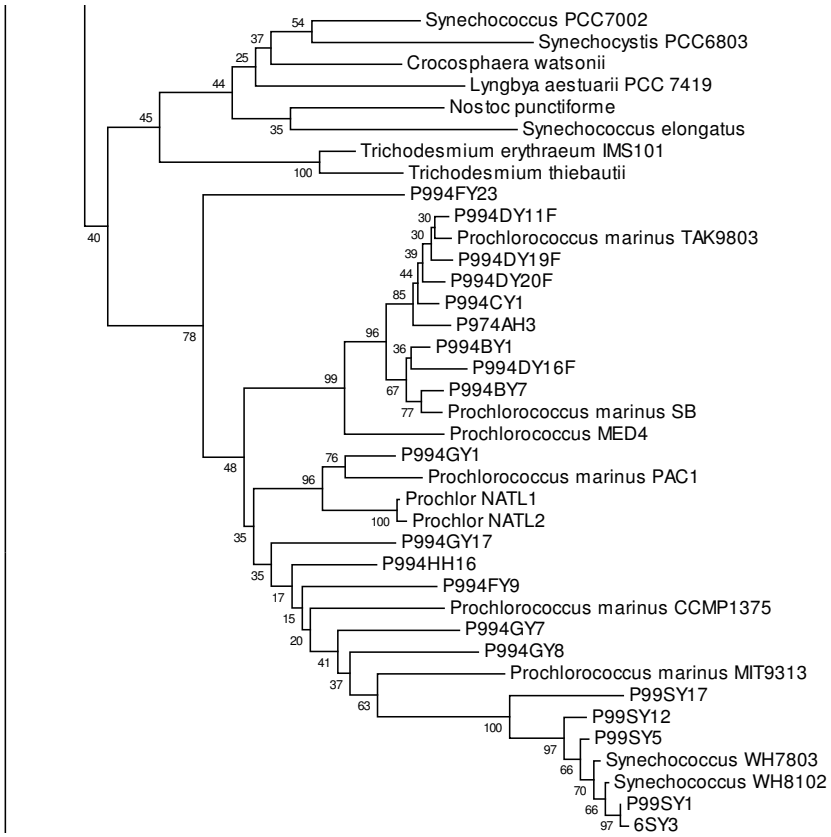


Figure A-5. Conensus tree of Form IA clone sequences obtained from the MRP in 1999 and 2001 along with their closest relatives.

Appendix A (Continued)

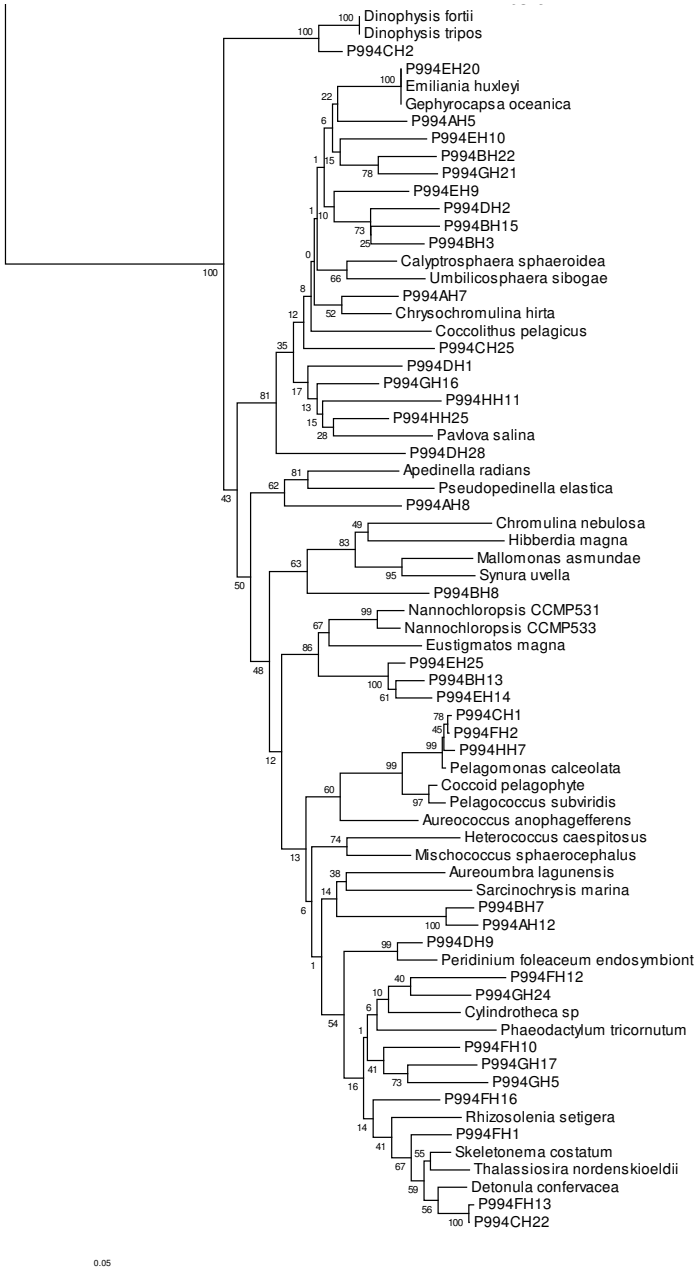


Figure A-6. Conensus tree of Form ID clone sequences obtained from the MRP in 1999 and 2001 along with their closest relatives.